UNIVERSITY OF LIVERPOOL

Understanding the Diffusion of Small Gases in Porous Organic Cages using Molecular Dynamics

Daniel Holden

Supervisor: Prof. Andrew Cooper

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy

September 2013

Abstract

The aim of this thesis was to accurately simulate the dynamic nature of known porous organic cage molecules, with a view to understanding the diffusion of different gases through their pore structures. Initial calculations showed that, due to their unique chemical structure, no 'off-the-shelf' force field (FF) was accurate enough to describe their dynamic motion. Therefore, a cage specific force field, (CSFF), was developed to be transferable across the first three cage systems, **CC1**-**CC3**.

CSFF was subsequently used to rationalise the 'on'/'off' porosity observed in two different polymorphs of **CC1**. A combination of computational simulations, including simulated surface area calculations, grand canonical Monte Carlo (GCMC) adsorption isotherms, and molecular dynamics (MD) simulations, for hydrogen and nitrogen in **CC1α** and **CC1β**, helped to confirm experimental results, as well as to provide further insight into why the polymorphism of **CC1** alters the porosity of the molecule.

In addition, CSFF was used to study the diffusion of a range of gases through crystalline **CC3**. Seven gases were chosen: hydrogen, nitrogen, carbon dioxide, methane, sulfur hexafluoride, krypton and xenon. A detailed understanding of the diffusivity within **CC3** was accomplished by combining MD simulations with new methodologies and techniques, for example analysis of the dynamic connectivity. This helped to rationalise why **CC3** showed good experimental uptake of gas, as well as highlighting potential separation capabilities.

In summary, the development of CSFF has made it possible to simulate the diffusion of small gases through porous organic imine cages, and it has been shown that this diffusion is dependent on the relative size of the gas to the cage window, assuming that there is a suitable diffusion pathway. Using MD simulations, we have unlocked phenomena such as gas selectivity, rare-event hopping and the diffusion of gases to regions previously thought inaccessible. This has aided the rationalisation of existing experimental observations, and is a significant step forward for *a priori* prediction of porous organic cage systems, and their properties. This work has also led to new experiments that were prompted by my simulations. Finally, a new way to visualise the connectivity of a system has been introduced. This is achieved by monitoring how the surface area evolves with respect to time, during a MD simulation. This suggests how the pore channels of various systems, previously thought too small for gas adsorption, are actually suitable candidates for separations.

Acknowledgements

I would like to express my sincere thanks to Prof. Andrew Cooper for giving me the opportunity to study a PhD within his group; as well as for all the help and guidance he has given me over the past few years. I am really grateful his input and direction, without which this work would not have been possible. Also, thanks for reading through my thesis; your corrections were invaluable.

I would also like to thank my second supervisor, Dr. Abbie Trewin, for all her assistance, especially as I got to grips with what I was meant to be doing! Also, credit is due to Dr. David Willock for his help in developing CSFF; without his patience and knowledge this would never have been completed.

A special thanks has to go to Dr. Kim Jelfs, for reading and correcting every page of my thesis at least twice – I hope it was worth all the effort! It would be fair to say that I think you're probably as glad as I am that it is finally finished!

It has also been a real pleasure to collaborate with Dr. Maciej Haranczyk. I really appreciate all the help he has given me with the dynamic connectivity plots, and look forward to continuing this work in the future.

In addition, I would like to thank fellow group members and friends, past and present, all who have contributed to my work in one form or another. Thank you, Tom H, Sam, Mark, Marc, Ed, Paul, Rob W, Rob C, Rob D, Ben and Shan.

I would also like to take this opportunity to thank my family for putting up with me over the past few months, so thanks Mum, Dad, Bex and Tom… I've finished now, lets go for a pint!

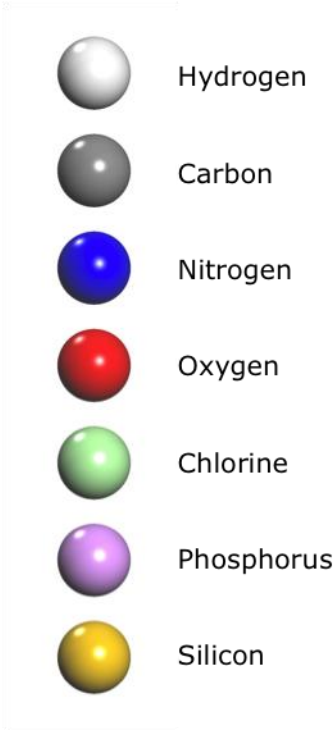
Finally, I want to thank and acknowledge God for the strength and commitment He has given me to persevere. He has helped me keep everything in perspective, and constantly reminded me why I work in the first place – to glorify Him. So thank you.

"I can do all this through Him who gives me strength."

Philippians 4:13

Atom colour key

Throughout this work, figures have the following colours for elements, unless otherwise indicated.



List of abbreviations

4TMSEBP	3,3′,4,4′-tetra(trimethylsilylethynyl)biphenyl
ASA	Accessible surface area
CB[n]	Cucurbit[n]urils
CBMC	Configurational-bias Monte Carlo
CMPs	Conjugated microporous polymers
COF	Covalent organic framework
CSA	Connolly surface area
CSD	Cambridge Structural Database
CSFF	Cage specific force field
CSP	Crystal structure prediction
DFT	Density functional theory
DoE	Department of Energy
$E_{bending}$	Three-body bending potentials
$E_{electrostatic}$	Electrostatic interactions
$E_{non\text{-}bonded}$	Sum of the intermolecular potentials
$E_{stretch}$	Sum of the two-body stretching potentials
$E_{torsion}$	Four-body torsional potentials
E_{total}	Sum total of the energy
E_{VDW}	Van der Waals interactions
FBM	Flexible body method
FF	Force field
GCMC	Grand canonical Monte Carlo
GGA	Generalised Gradient Approximation
HF	Hartree-Fock

IAST	Ideal adsorbed solution theory
IUPAC	International Union of Pure and Applied Chemistry
LDA	Local Density Approximation
MC	Monte Carlo
MD	Molecular dynamics
MOF	Metal–organic framework
NASA	Non-accessible surface area
PBC	Periodic boundary condition
PCFF	Polymer consistent force field
PES	Potential energy surfaces
PIM	Polymers of intrinsic microporosity
PLD	Pore limiting diameter
PLE	Pore limiting envelope
PNC	Phthalocyanine metal nanoporous crystal
RBM	Rigid body method
RMS	Root mean squared
SA_{BET}	Brunauer–Emmett–Teller surface area
SA_{LANG}	Langmuir surface area
SSA	Solvent surface area
STP	Standard temperature and pressure
TBC4	P-tert-butylcalix[4]arene
TPP	Tris-o-phenylenedioxycyclotriphosphazene
TTBI	Triptycene trisbenzimidazolone
UFF	Universal force field
ZIF	Zeolitic imidazolate framework

Contents

Chapter 1

Introduction

1.1. Microporous materials

Porous materials are subdivided into different categories depending on their pore dimensions, as defined by the International Union of Pure and Applied Chemistry (IUPAC).[1] Pore sizes range: 50-1000 nm (macropores), 2.0-50.0 (mesopores), and 0.2-2.0 nm (micropores). This work focuses on materials found within the microporous region. Microporous materials are important in a wide range of areas, including heterogeneous catalysis, gas storage, and molecular separation.[2-6] It is possible for microporous materials to exhibit high surface areas, and to adsorb gases at lower pressures.[7,8] These materials are attractive to chemists because the size of their pores is comparable to the size of small molecules.[9,10] Most microporous solids are extended networks, such as zeolites, activated carbons, metal–organic frameworks (MOFs)[11], covalent organic frameworks (COFs),[4] or polymer networks,[12,13] however a large amount of research has also focused on porous molecules. Examples of porous molecules include calix[n]arenes,[14] dipeptides,[15] cucubit[n]urils,[16] and porous organic cages.[17,18] These molecules are not interconnected by strong covalent or coordination bonds, and this gives scope for dynamic motion and flexibility. It has been known for some time that molecular crystals can possess significant local flexibility that can facilitate guest diffusion, even in materials which are formally non-porous.[19] However, measuring porosity in these molecules can be subjective, therefore a framework of definitions is required. These are covered in the next section.

1.2. Porosity and crystallinity

The propensity to adsorb guests into a connected network of permanent void volume can be defined as physical porosity.[20] This definition is useful and easily applicable for materials where their pore dimensions are much larger than the sorbates, for example mesoporous materials, however it becomes harder to define whether a system is conventionally porous, or exhibits 'porosity without pores' with formally disconnected voids, when the size of the pore apertures are close to the size of the guest molecules.[2] In such cases, a static, rigid view of the structure will not adequately represent the potential guest sorption properties. It is clear that this phenomenon would only be seen if the flexibility of the system is taken into account. This is because the definition of pore 'connectivity' from static crystal structures, based on probe spheres of various diameters, will miss transient channel openings, resulting from thermal fluctuations.

The term 'porosity' can have a variety of meanings. For clarity in this thesis the same terminology described previously by Barbour is used throughout.[2] This states that two key features are important:

1. The term 'porous' should apply to a specific host system, and not just a collection of amorphous material. The structure of the host should be well defined, with the pore connectivity known.[21]

2. As 'permeable' and 'porous' are synonymous, permeability should be demonstrated for a system to be classed as porous. This means that guests should be able to freely diffuse through the pore network.[21]

The porosity of porous molecular crystals is dependent on the constituent molecule's shape, and can be attributed as being either from the intrinsic or extrinsic regions within the host system. These are defined as:

- **Intrinsic porosity**: porosity that is generated from the structure of the discrete molecules when viewed in isolation; pre-fabricated voids being contained within each molecule. This is applicable for any bowl-shaped or cage like materials, for example calix[n]arenes or porous organic cages.[14,17,22]

- **Extrinsic porosity**: porosity generated between the discrete molecules, and is generated as a result of the inefficient packing of the molecular subunits.[20,22] The molecule itself needs not contain any 'pre-fabricated' void cavities.[18]

It is possible for a molecular system to exhibit both intrinsic and extrinsic porosity. Molecules can also crystallise in a number of different packing motifs, and this is known as polymorphism.[23] This can be controlled by a number of factors, including alteration in the reaction conditions, for example changing the solvent, and the subsequent generation of different packing arrangements; this has a direct influence on the properties of the material.[2]

It is also possible to amorphise porous molecules, either by design, using methods like freeze-drying,[24] or when crystallisation is prohibited by the shape of the molecules.[25] This can lead to enhanced properties, for example increased surface area, due to additional extrinsic porosity, but may also reduce the surface area by packing the molecules in such a way that pore connectivity is lost.[26] A scheme highlighting how porous molecules can pack is displayed in Figure 1.1.

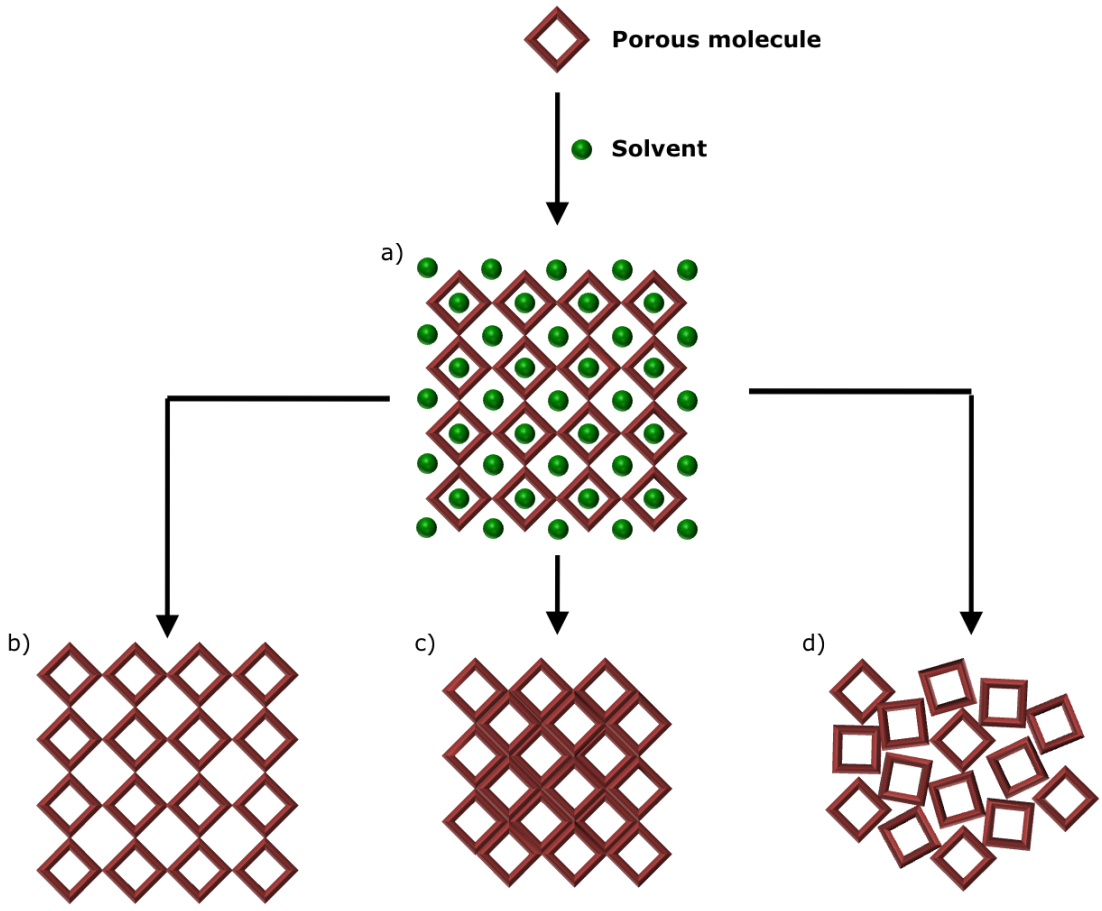


Figure 1.1 – Potential solid state packing arrangements for an intrinsically porous molecular material;[27] a) The solvated crystal structure of a porous molecule can be desolvated in a number of ways, to form, b) the ideal structure, where the loss of solvent has not affected the packing motif, thus leaving both intrinsic and extrinsic porosity. However, it is more likely to rearrange to form a more dense structure to minimise free space, c), where only intrinsic porosity is available. Alternatively, d) an amorphous product could form.

1.3. Porous networks

There are a variety of different porous networks, including zeolites, metal-organic frameworks (MOFs), covalent organic frameworks (COFs), conjugated microporous polymers (CMPs), and polymers of intrinsic microporosity (PIMs). This section briefly describes the discovery and synthesis of these materials, highlighting potential applications. Examples of the structures of porous networks can be found in Figure 1.2.

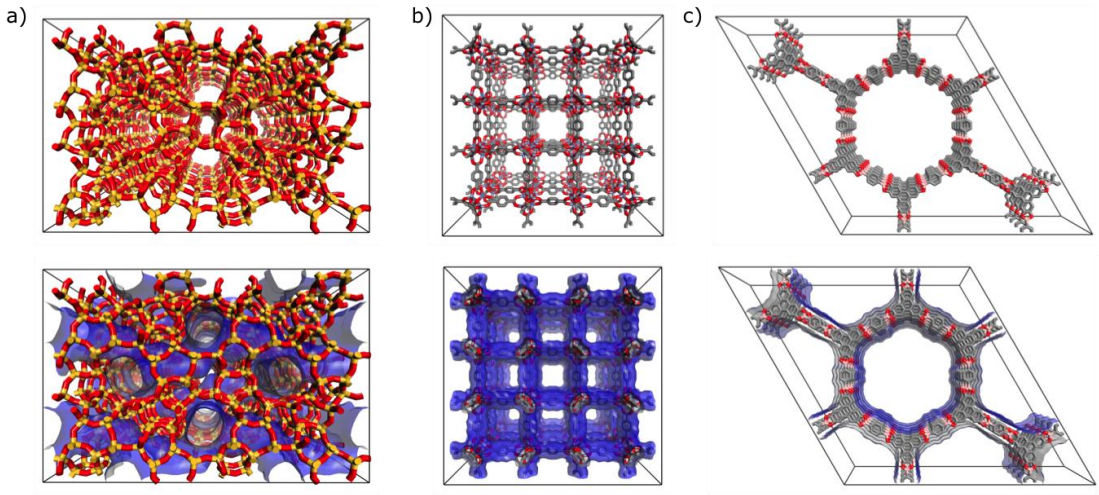


Figure 1.2 – Figure showing the crystalline network of, a) a zeolite (MFI framework topology), b) MOF-5, and c) COF-5. The Connolly surface area for each are shown below in blue and these were generated using a probe radius of 1.55 Å. Hydrogens have been omitted for clarity.

Zeolites were first discovered by Cronstedt in 1756,[28] and are both naturally occurring and synthetically produced. Zeolites are crystalline aluminosilicates with a fully cross-linked framework consisting of corner-sharing SiO_4 and AlO_4 tetrahedra.[29] They are crystalline solids that exhibit three-dimensional structures with pore cavities ranging from 5-15 Å in diameter, and templates are often used to control the framework topology, and thus pore size. Many different structures are possible, as the tetrahedral building blocks can be linked in a vast array of combinations. This has led to a wide range of different zeolite topologies. Each topology is identified by a three letter "Framework Type Code", such as MFI in Figure 1.2a, and these all have their own regular array of channels and pore cavities.[3] In all, over 130 different framework structures are known.[30] These are ideal materials for heterogeneous catalysis, and are more specifically used for cracking catalysis; this is a multi-billion dollar industry.[3,31] As well as this, other applications include ion exchange,[3] molecular sieving,[32-34] and in laundry detergents.[3,35]

MOFs are another category of porous networks. The synthesis and characterisation of these is one of the most rapidly developing areas of chemical science, and this is due to their synthetic versatility, large pore size, high apparent surface areas, measured selective uptake of small molecules, and optical or magnetic responses to the inclusion of guests.[36,37] MOFs have an extended crystalline structure which consists predominantly of metal clusters and organic linkers.[38] Strong covalent bonds between these ligands and metal clusters form one-, two-, or three- dimensional structures.[39] Due to the large variety of

metal clusters and organic linkers available, the number of potential MOFs is huge. This led to Yaghi and O'Keeffe proposing 'reticular chemistry' – the design and construction of MOFs by the use of secondary building units (SBUs).[11] Reticular synthesis is the process of assembling rigid molecular building blocks into pre-arranged ordered structure and these are held together by strong covalent bonds.[11] Throughout this process, the structural integrity and rigidity of the building blocks remains constant.[11] The metal clusters are linked together by organic linker to produce the frameworks.[40] Assuming that the appropriate synthetic conditions are used, this strategy makes it possible to design MOFs with well-defined pore shapes, sizes, and functionalities, simply by altering the organic linker. This means that, in theory, these materials can be tuned so that they possess the desired properties.

In addition, MOFs are good candidates for gas separation. They have shown separation of hydrogen from a number of other gases,[41] selective adsorption of carbon dioxide over nitrogen,[42] xylene from other alkane isomers,[43,44] and noble gas separation.[45] A disadvantage of MOFs is that, in some cases, their pore structure collapses upon desolvation,[46] and many exhibit poor thermal or hydrolytic stability.[47] The major difference between MOFs and COFs is that COFs are composed of purely organic atoms.[48] Boron, carbon, oxygen and silicon atoms create strong covalent bonds to form either two-dimensional layered structures, or three-dimensional networks.[49,50] These have rigid structures and low densities, creating permanent porosity, and as a result these can have surface areas that surpass those seen in zeolites. COFs can in some cases exhibit exceptional thermal stabilities of temperatures up to 600 °C.[4] The first two-dimensional COF synthesised, COF-1, was reported by Côté *et al.* in 2005 *via* a dehydration reaction of 1,4-benzenediboronic acid (BDBA); three boronic acid molecules formed a planar six-membered boroxine ring, as three water molecules were eliminated – subsequently these rings produced a staggered layered framework.[4]

The first reported three-dimensional COFs were reported in 2007 by El-Kaderi *et al*.[51] These were constructed as crystalline solids *via* the self-condensation reaction of tetrahedral tetra(4-dihydroxyborylphenyl)methane (TBPM), or the silicon analogue tetra(4-dihydroxyborylphenyl)silane (TBPS), and by co-condensation with triangular 2,3,6,7,10,11-hexahydroxytriphenylene (HHTP).[51] As these consist of only strong covalent bonds, they also exhibit high thermal stabilities (up to 500 °C).[51] These three-dimensional COFs have very low crystallographic densities, leading to high surface areas; for example COF-108 has a density of 0.17 g cm^{-3} while COF-103 displays a Brunauer-Emmett_teller

surface area (SA_{BET}) of 4210 m^2 g^{-1}.[52] In general, COFs have lower densities than MOFs for a given topology, due to the absence of any heavy metal atoms, and this increases the mass percentage of guest molecules that can be adsorbed into these materials.[50] This has led to high uptake of gases, with Furukawa *et al.* reporting that COF-1 displayed a sorption capacity of 2.5 mmol g^{-1} (35 bar, 298 K); this exceeded the methane storage target set by the Department of Energy (DoE).[53]

Two examples of amorphous porous networks are CMPs and PIMs. Unlike crystalline framework materials, these are formed under kinetic control and are consequently amorphous in nature, possessing no long-range order.[54] In 2007, Jiang *et al.* reported the first CMP, and this was synthesised *via* a Sonogashira–Hagihara coupling reaction of aromatic alkynes with aromatic halides, with the resultant material exhibiting a SA_{BET} of around 834 m^2 g^{-1}.[54] CMPs typically have good chemical and thermal stability and their physical properties, such as surface area, micropore size and gas uptake, can be controlled by varying the length of the molecular struts[55] – this has an impact on both pore size and micropore volume. Small molecular struts maximise the micropore volume, with small pores are observed, while long molecular struts maximise the pore size, but reduce the micropore volume.[55] Atomic simulations have been used to suggest that by using shorter strut lengths the flexibility between the struts and adjacent connected benzene nodes is reduced.[55]

Recently, a wide range of applications for CMPs has been investigated. Jiang *et al.* have shown that it is possible to incorporate metals, such as rhenium, rhodium and iridium, into CMPs for use as heterogeneous catalysts.[56] Light harvesting has also been reported by Chen *et al.* in polyphenylene CMPs, and these materials have a pore size of 15.6 Å, with a SA_{BET} of 1083 m^2 g^{-1}.[57] The same group have also designed metalloporphyrin CMPs that exhibit a SA_{BET} of 1270 m^2 g^{-1}, and these can also be used as heterogeneous catalysis for the oxidation of sulfides with a selectivity of 99%, and a 98% conversion rate.[58]

PIMs are porous polymers initially reported by Mckeown and Budd in 2002.[59] This study demonstrated that it was possible to derive microporous materials directly from soluble polymers that orientate in a random manner to prevent efficient packing in the solid state.[60] Originally networks based on porphyrins[59] and phthalocyanines,[61] second-generation PIMs are soluble polymers that have a rigid, contorted molecular structure.[59] These can be prepared as insoluble networks or as a soluble linear polymer.[62] One major advantage of PIMs is that

they can be cast as microporous films, with good mechanical properties.[22] This has led to applications in gas separation where high gas flux across thin films is desirable;[22] high permeability and gas separation of oxygen from nitrogen has been observed in PIM-1.[62,63]

PIMs can also exhibit large surface areas. Porphyrin-based PIMs have been reported to have a SA_{BET} of up to 1000 m^2 g^{-1}, while phthalocyanines derivatives, show a SA_{BET} of anything between 450-950 m^2 g^{-1}.[59,61] In 2006, Ghanem *et al.* synthesised a triptycene-based PIM that displayed an enhanced surface area (1065 m^2 g^{-1}), with a reversible sorption of hydrogen by mass of 1.65% (1 bar, 77 K) and 2.71% (10 bar, 77 K).[64] It has also been shown by Budd *et al.* that it is possible to synthesise PIMs containing catalyst supports.[65] This ability to incorporate a varied range of functional units into PIMs has led to applications in chemoselective adsorption as well as heterogeneous catalysis.[65]

1.4. Porous molecules

Porous molecules differ from porous networks, as each system is comprised of discrete subunits, as opposed to a fixed framework. Inspiration for these has come from a variety of different sources, including the mimicry of nature, the use of biological molecules, and inefficient packing to generate extrinsic porosity. Different examples of porous molecules can be found in Figure 1.3.

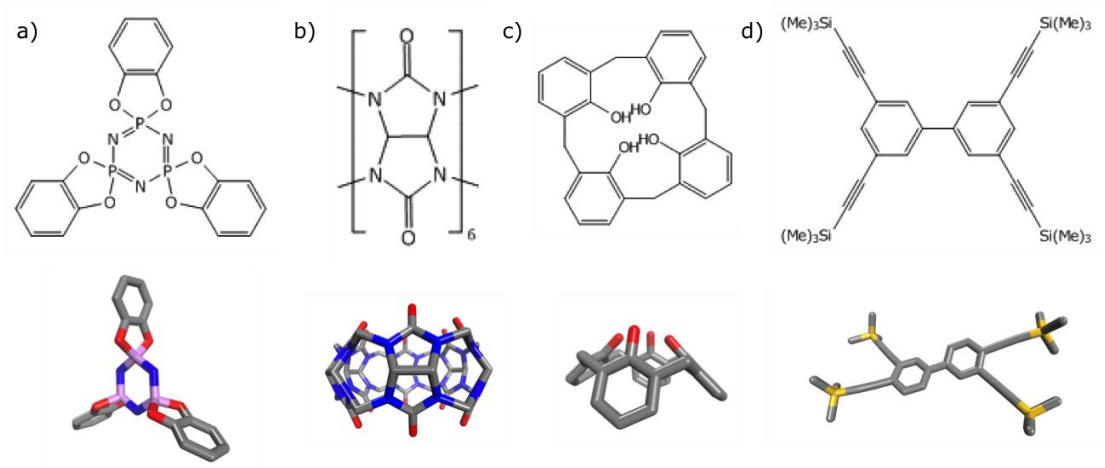


Figure 1.3 – Schematic representation of four different porous molecules: a) tris-o-phenylenedioxycyclotriphosphazene (TPP), b) cucubit[6]uril, c) calix[4]arene, and d) 3,3',4,4'-tetra(trimethylsilylethynyl)biphenyl (4TMSEBP). Hydrogens have been omitted for clarity.

Dubbed as an 'organic zeolite' by Barrer *et al.* in 1976, Dianin's compound was shown to absorb a number of gases, including the noble gases argon, krypton and xenon.[66] This was reported as the first discrete porous molecular solid. Six Dianin molecules are held together *via* hydrogen bonds and one-dimensional

hourglass channel are formed between them; this is only 2.5 Å is diameter, meaning it is not very porous.[67]

Calix[n]arenes were also among of the first reported 'porous molecules', and these were introduced in 1983 by Gutsche *et al*.[70] These demonstrated weak, non-covalent interactions between each discrete unit, mimicking natural recognition processes.[14,68,69,70] Calix[n]arenes are a versatile class of macrocyclic compounds that have a basket-shaped conformation, which promotes accommodation of guest molecules.[14] They can also be used as both host materials and platforms for the synthesis of specifically designed receptors.[14,68,69] They are traditionally formed from the hydroxyalkylation of a phenol with an aldehyde. Calix[4]arene is the simplest known calix[n]arene, and the structure is stabilised by a cyclic array of hydrogen bonds between the adjacent phenolic alcohol groups on the lower rim.[71] Host-guest systems of calix[4]arene and gases like methane or freon molecules can form, with the gases filling the interstitial voids of the system.[71] Many other calix[n]arene derivatives have since been synthesised by adapting the functionality of either the upper of lower rim of the molecule.[12,72]

Calix[n]arenes can also exhibit polymorphism. Atwood *et al.* have shown that it is possible to form a low density polymorph of p-tert-butylcalix[4]arene (TBC4) *via* sublimation;[19] this compound underwent a single-crystal-to-single-crystal phase transition when exposed to a guest. Even though the porosity was relatively low, guest transport was observed until a thermally stable structure was obtained.[19] This single-crystal-to-single-crystal transformation required the discrete molecules to undergo a large positional and orientational rearrangement, and this was possible due to only weak van der Waals interactions between the molecules being present in the system. Uptake of vinyl bromide initiated this transformation, such that each cavity of the TBC4 encompassed one vinyl bromide, resulting in a 1:1 host-guest complex.[19] Cooperative diffusion of the vinyl bromide in TBC4 was achieved by a shift in the layers. Experimental gas isotherms showed that at 1 atm, 80% of the cavities were occupied by carbon dioxide, rising to 100% at 3 atm. In contrast to this, hydrogen sorption was not observed, even at 7 atm. This meant that TBC4 is an ideal candidate for the separation of carbon dioxide from hydrogen.[69]

Another molecule inspired by the natural world was introduced by Mock *et al.*, when they named the macrocyclic methylene-bridged glycoluril hexamer "cucurbituril" down to its resemblance to the cucurbitaceae family of fruits

(pumpkins).[16] Cucurbit[n]urils (CB[n]) are macrocyclic molecules made up of *n* (where commonly *n* = 5,6,7, 8, or 10) glycoluril repeat units, and are formed by the condensation of cheap starting materials, for example glycoluril with formaldehyde.[73] These molecules also display a high thermal and chemical stability.[74] As the cavity within CB[n] is non-polar, it is a good candidate for the binding of hydrophobic guests.[73]

Lim *et al.* showed that CB[6] demonstrated a high adsorption capacity for acetylene, with an acetylene uptake of 4.2 mol per mole of CB[6] reported.[75] Scherman *et al.* also reported in 2008 that CB[6] could be dissolved *via* the encapsulation of an imidazolium-based ionic liquid, and that the dissolution ability could be tuned by altering the structure of the imidazolium used.[76] The formation of a 1:2 inclusion complex with CB[6] facilitated a new type of self-assembly. More recently, Atwood *et al.* showed that CB[7] exhibited an even higher uptake of carbon dioxide than CB[6], with a carbon dioxide uptake of 2.3 mmol g^{-1} (1 bar, 297 K). However, the uptake of nitrogen and methane was surprising low, and this made CB[7] an ideal material for selective carbon dioxide adsorption.[77] In addition, Kim *et al.* reported in 2013 that the recrystallisation of CB[6] from hydrochloric acid gave a porous material with a one-dimensional channel. This had a pore volume of 0.13 cm^3 g^{-1} and a SA_{BET} of 210 m^2 g^{-1}.[74] They also carried out sorption isotherms for CB[6] for a number of gases, including and carbon monoxide, carbon dioxide and methane, and this showed unusually high levels of carbon dioxide uptake of 2.2 mmol g^{-1} (1 bar, 298 K), and at the time this was much higher than any known organic molecular porous materials.[74] This can be explained by looking at the carbon dioxide adsorption sites, as there are two per CB[6] – one in the one-dimensional channel, and one in the CB[6] cavity.[74]

Noria is another type of porous molecule. Atwood *et al.* showed in 2009 that this cyclic, waterwheel-like oligomer was formed in the amorphous state, and the structure was determined by X-ray diffraction and scanning electron microscopy.[78] Although the SA_{BET} for nitrogen was measured to be just 40 m^2 g^{-1}, upon the uptake of carbon dioxide, the SA_{BET} was increased to 250-280 m^2 g^{-1}. Originally, it was thought that this was facilitated by the large internal cavity within Noria - this has a volume of around 160 $Å^3$, with a pore diameter of approximately 5 Å.[22] However, this internal volume is too small to explain the increased uptake of carbon dioxide, thus implying that both intrinsic and extrinsic porosity must be found within the sample. The internal volume can be further increased to 350 $Å^3$ by attaching bulky tert-butyloxycarbonyl or methacrylate groups to the edge of the Noria structure.

It is also possible to use biological molecules as alternative building blocks for the generation of porous materials. These offer a variety of previously overlooked structures, and have targeted interactions built-in that, in principle, could be exploited, and are also biocompatible and biodegradable.[79] Recently, permanent porosity has been recognised in dipeptide architectures.[15] Dipeptide crystals contain channels of various sizes, and these show selective absorption of methane, carbon dioxide and hydrogen.[80] Due to the high number of peptide monomers, in combination with their sequential arrangement, the synthesis of robust porous materials has been accomplished.[81] Here, the size, shape and helicity of the channels can be tuned.[81]

Four crystalline dipeptides, L-alanyl-L-valine (AV), L-valyl-L-alanine (VA), L-isoleucyl-L-valine (IV) and L-valyl-L-isoleucine (VI), can be assembled in a hexagonal manner to form one-dimensional channels, with hydrogen bonds providing stabilisation. Each of these form hydrophobic channels of varying size: AV = 5.0 Å, VA = 4.7 Å, IV = 3.9 Å and VI = 3.7 Å.[80] Looking specifically at the storage capacity in dipeptides VA and AV, Sozzani *et al.* reported that VA exhibited uptakes of 4.1 mmol g^{-1} carbon dioxide, and 2.2 mmol g^{-1} methane, while AV displayed uptakes of of 3.5 mmol g^{-1} carbon dioxide, and 1.6 mmol g^{-1} methane (1 atm, 195 K). Consequently, their CO_2/CH_4 selectivity is around 2.5. However, uptake in dipeptides is not just limited to carbon dioxide and methane, as Soldatov *et al.* has also reported that both AV and VA have a high affinity for the adsorption of xenon.[81]

Haemoproteins are another biological molecule that have been imitated by synthetic chemists for the synthesis of porous molecules, most notably phthalocyanine metal nanoporous crystals (PNCs).[82] Haemoproteins are involved in facilitating the storage and transport of oxygen (haemoglobin and myoglobin), oxidative catalysis in cytochrome P450, and the sensing of carbon monoxide in guanylyl cyclase.[82] In 2010, Bezzu *et al.* recognised that the crystal engineering of nanoporous structures using these 'heme' like motifs had not been exploited, and thus reported an iron phthalocyanine derivative that formed millimetre-scale molecular crystals.[82] These contained large interconnected voids (around 80 $Å^3$), formed by the cubic assembly of six phthalocyanines. These exhibit type 1 nitrogen isotherms, and possess high surface areas in the range of 850 to 1000 $m^2\ g^{-1}$, with pore volumes of around 0.46 ml g^{-1}.[82] One advantage of PNCs is that they are compatible with aqueous media. Rapid ligand exchange can also be achieved within these nanoporous crystals by single-crystal–to–single-crystal transformations.[82] PNCs nanoporous structure ensure rapid access of ligands to

the axial binding sites, even by simple adsorption from the gas phase. To prevent structural collapse during desolvation, all these axial ligands can be replaced *via* an exchange reaction.

Another approach to uncover porous molecules has been to mine existing databases of materials. This method led to the unearthing of 3,3',4,4'-tetra(trimethylsilylethynyl)biphenyl (4TMSEBP) from the Cambridge Structural Database (CSD) by Msayib *et al.* in 2009.[83] Originally discovered in 1983 by Larsson, this was uncovered by screening the database for molecules possessing specific properties; first molecules with low crystallographic densities were chosen, and subsequently any molecules that contained rigid aromatic rings were selected, as these induced stability as well as enhancing the potential of microporosity. One such molecule was 4TMSEBP. A striking feature of this molecule was its resemblance in structure to zeolites. 4TMSEBP's crystal structure is stable upon desolvation, and exhibited internal voids with diameters of 11 Å; this was produced by the intersection of three-dimensional orthogonal channels, all around 4 Å in diameter.[22] 4TMSEBP exhibited an uptake of 3.9 mmol g^{-1} of hydrogen (10 bar, 77 K), and a nitrogen SA_{BET} of 278 m^2 g^{-1}.

Originally discovered by Allock *et al.,* tris-o-phenylenedioxycyclotriphosphazene (TPP) also exhibits extrinsic porosity, and has a 'paddle-wheel' molecular structure.[84] This has a phosphazene (P_3N_3) core, with three orthogonal catechol rings, and crystallises to form a structure with one-dimensional pores; this channel is around 4.5 Å in diameter.[22,83] Sozzani *et al.* also reported that the encapsulated solvent molecules could be removed without loss of the porous structure, and this left 25% free volume within the crystal structure.[85,86] Uptake of various of gases, for example argon, nitrogen, oxygen, hydrogen methane, carbon dioxide, and iodine have been seen in TPP.[83,85-87] The measured nitrogen SA_{BET} for TPP was calculated to be 240 m^2 g^{-1}, while the carbon dioxide uptake was determined to be at around 4 wt% (800 mbar, 298 K). In addition, xenon showed a remarkable affinity to the trigonal cavity found within TPP, and occupied around 90% of these sites, as the pressure was increased to 100 kPa.[88] The stability of xenon in the TPP cavities is explained by the favourable interaction of xenon with the π-electron rich environment of the phenylenedioxy-wall.[89] As TPP contains no internal pores, there is no intrinsic porosity, meaning that any surface area is entirely extrinsic. This is because of the formation of hexagonal channels in the crystal structure.

Another example of an extrinsically porous molecule was reported by Mastalerz and coworkers in 2012; triptycene trisbenzimidazolone (TTBI).[90] This undergoes self-assembly *via* the formation of hydrogen bonds, resulting in the formation of porous one-dimensional channels. The first channel is almost cylindrical in nature, and has an average diameter of 14.5 Å. The second channel is more slit-like in nature and these are located between the cylindrical channels. TTBI absorbs around 33.7 mmol g^{-1} of nitrogen at 77 K, with a partial pressure of 0.95, while the measured SA_{BET} was 2796 m^2 g^{-1}, and the Langmuir surface area (SA_{LANG}) is 3020 m^2 g^{-1} – this is one of the highest for all organic compounds (both those with intrinsic or extrinsic porousity).[90] In addition, TTBI shows high selectivity for carbon dioxide over methane, 3.6 mmol g^{-1} and 0.9 mmol g^{-1} respectively (1 bar, 273 K). The hydrogen uptake was also high at 11 mmol g^{-1} (71 bar, 77 K).

1.5. Porous imine cages

As well as the porous molecules mentioned above, a variety imine cage molecules have been synthesised, and these exhibit porosity to a number of small gases. In 2006, Warmuth *et al.* used dynamic covalent chemistry to synthesise a number of molecular nanocages, *via* the condensation reactions of tetraformylcavitands with a number of different diamines in the presence of a trifluoroacetic acid catalyst.[91] This was a simple one-pot synthesis and by altering the synthetic conditions, for example the solvent, different shaped nanocages could be created;[92] this included the formation of tetrahedral, octahedral and square-antiprismatic nanocages. The [2+4] tetrahedral nanocages were synthesised using two tetraformylcavitands and four diamine linkers (either 1,3-diaminopropane or 1,4-diaminobutane), while the [6+12] nanocages were synthesised using six tetraformylcavitands and twelve diamine linkers (1,2-ethylenediamine linkers). The largest [8+16] square-antiprismatic nanaocages were synthesised using eight cavitands and sixteen diamines.[93] As these nanocages possess large internal cavities, the aim was to use these imine molecules to encapsulate different guest molecules, however no crystal structures or porosity have been reported.[94]

In 2008, Mastalerz reported some shape persistent endo-functionalised nanocage compounds.[95] Like Warmuth, these were synthesised using imine condensation *via* a one-pot synthesis. The dynamic nature of these reversible reactions prevented the formation of the kinetic product. The first of these endo-functionalised cages was synthesised using four triamine corner units with six salicylic dialdehyde linkers. In 2010 their porosity was evaluated, and based on the nitrogen isotherms, with SA_{BET} of 1375 m^2 g^{-1} and a SA_{LANG} of 1566 m^2 g^{-1}

calculated.[96] In addition, uptake of 2.1 mmol g^{-1} carbon dioxide and 0.6 mmol g^{-1} methane was observed (1 bar, 273 K), and this corresponded to 9.4 wt% carbon dioxide, 0.94 wt% methane. This highlighted a good selectivity of carbon dioxide over methane (10:1).[96]

It has also been shown that it is possible to decorate the periphery of these salicylbisimine cages. This has led to the generation of a family of shape-persistent, periphery-substituted cage compounds, that exhibit varying levels of permanent porosity.[97] Their surface area was dependent on whether they were synthesised in a crystalline or amorphous state.[97] When crystalline, the SA_{BET} of the functionalised salicylbisimine cages was heavily dependent on the bulkiness of the peripheral group – by switching the peripheral group from a methyl group, to 3-ethylpentyl groups, and then to a trityl group, the SA_{BET} decreased from 1291 $m^2 g^{-1}$, to 309 $m^2 g^{-1}$, to 22 $m^2 g^{-1}$. The most promising peripheral group was tert-butyl, as this displayed a SA_{BET} of 2071 $m^2 g^{-1}$.[97] However, when packed in an amorphous manner, the SA_{BET} ranged from 690 to 727 $m^2 g^{-1}$.

Mastalerz *et al.* subsequently developed a synthetic approach that controlled the shape and size of these salicylbisimine cages, and this was achieved by mixing the reactants in a binary solvent mixture.[98] The surface area of salicylbisimine cages was higher when packed in an amorphous manner, with the SA_{BET} increasing from 309 $m^2 g^{-1}$ for the crystalline cages, to 700 $m^2 g^{-1}$ when packed in the amorphous state, and this may be because of additional extrinsic porosity.[98] As well as these [4+6] salicylbisimine cages, Mastalerz *et al.* demonstrated in 2012 that it was possible to generate [2+3] derivatives, using the same triamine corner unit as before, with longer salicylic dialdehydes, *via* imine condensation.[99] It was possible to increase the length of the [2+3] salicylbisimine cage by elongating the salicylic dialdehyde used. This affected the surface area – the shortest organic linker reported a SA_{BET} of 744 $m^2 g^{-1}$, while the longest organic linker showed a minimal SA_{BET} of only 30 $m^2 g^{-1}$.

In 2010, Zhang *et al.* also synthesised some [2+3] porous organic cages, using the same cycloimination reactions, and these packed such that a small one dimensional channel was seen to run through the crystal structure.[100] As the internal void was only small, low gas uptake was observed – 0.20 mmol g^{-1} carbon dioxide and 0.061 $cm^3 g^{-1}$ nitrogen (1 bar, 293 K). This corresponds to a high selectivity of 73:1. Zhang *et al.* then synthesised other imine cages by using readily available starting materials, while demonstrating that these cages were both well-defined and tunable.[101] As well as this, the cages were covalently cross-

linked together to form a cage framework, and these cage framework materials were reported to display extraordinarily high selectivity of carbon dioxide over nitrogen (138/1), with stability observed up to a temperature of 710 K.[101] However, the overall carbon dioxide capacity for both these systems was lower than other published organic molecules,[17] and Jiang *et al.* pointed out that although both these selectivities are high, the uptake of gases is extremely low and therefore the storage capacity should also be considered.[102]

1.6. Porous organic cages in the Cooper group

In 2009 Cooper and co-workers reported three tetrahedral porous organic cages, **CC1**, **CC2** and **CC3**. These were synthesised using a one-step [4+6] cycloimination reaction, of four 1,3,5-triformylbenzene units with six diamine linkers; alteration of the diamine linker dictated which cage was formed. 1,2-ethylenediamine was used to construct **CC1**, and this generated six un-functionalised vertices. 1,2-propylenediamine was used to produce **CC2**, and this decorated the outside the cage with a single methyl group on each vertex; two sites per vertex are available for these methyl groups. Finally, **CC3** was formed using (*R,R*)-1,2-diaminocyclohexane, and this meant that each **CC3** vertex contained a bulky cyclohexyl group. The reaction scheme for these cages is shown in Figure 1.4. **CC1**-**CC3** all have an intrinsic, covalently prefabricated molecular void. This thesis concerns the development of a FF, and the subsequent modelling of two **CC1** polymorphs, **CC1α** and **CC1β**, as well as diffusion of small gases through **CC3**.

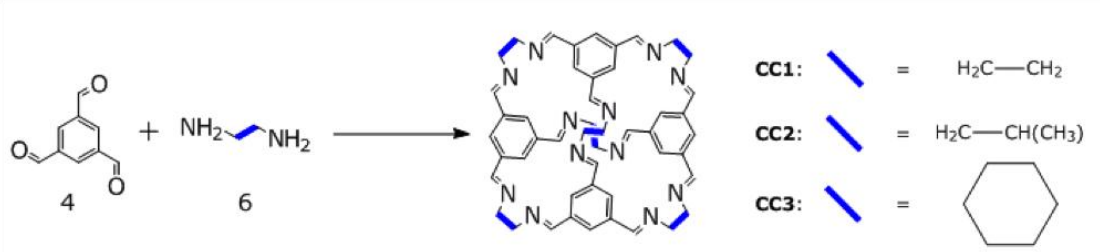


Figure 1.4 – Cycloimination reaction of four aldehydes with six diamines, to synthesise an imine cage.

Varying the diamine has a large impact on the packing motif of the cages, with the porosity observed in these crystalline materials a result of inefficient crystal packing combined with the pre-fabricated cage cavities. An overall scheme displaying the synthesis of **CC1**-**CC3** is shown in Figure 1.5.

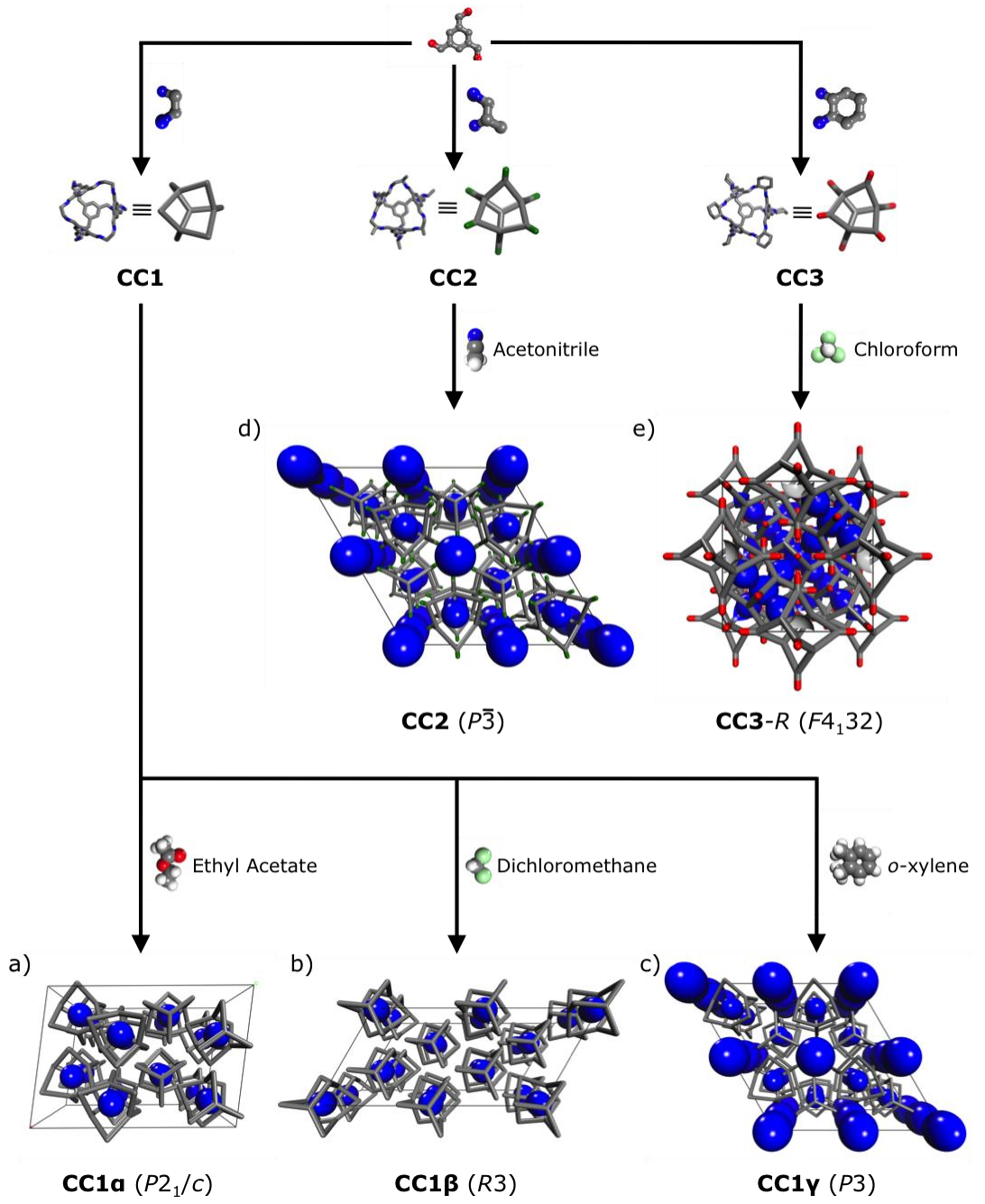


Figure 1.5 – Schematic pathway showing the reactants and solvents required to crystallise a) **CC1α**, b) **CC1β**, c) **CC1γ**, d) **CC2**, and e) **CC3**. Reaction scheme for cage synthesis is shown in Figure 1.4. Schematic representations of the cages have been used in the crystal structures, and the pore topology has been calculated using a nitrogen probe.

CC1 can be recrystallised using different solvents to give three different polymorphic structures,[103] **CC1α**, **CC1β**, and **CC1γ**. **CC1α** was crystallised in the presence of ethyl acetate and the measured uptake showed that it was non-porous to both hydrogen and nitrogen,[104] as the discrete units pack in a window-to-arene configuration throughout the crystal structure, thus restricting the pore topology to display only isolated voids (Figure 1.5a). The crystallisation of **CC1β**

in the presence of dichloromethane altered the crystal packing motif of the discrete units, such that experimental uptake of **CC1β** is selectively porous to hydrogen. **CC1α** and **CC1β** can also be interconverted in a reversible manner in the solid state, by exposure to organic triggers, thus allowing the porosity to nitrogen to be switched 'on' and 'off'; however, a conformational change of the cage molecules was required from tetrahedral to C_3 molecular symmetry.[103] The experimental uptakes of **CC1α** and **CC1β** are shown in Figure 1.6.

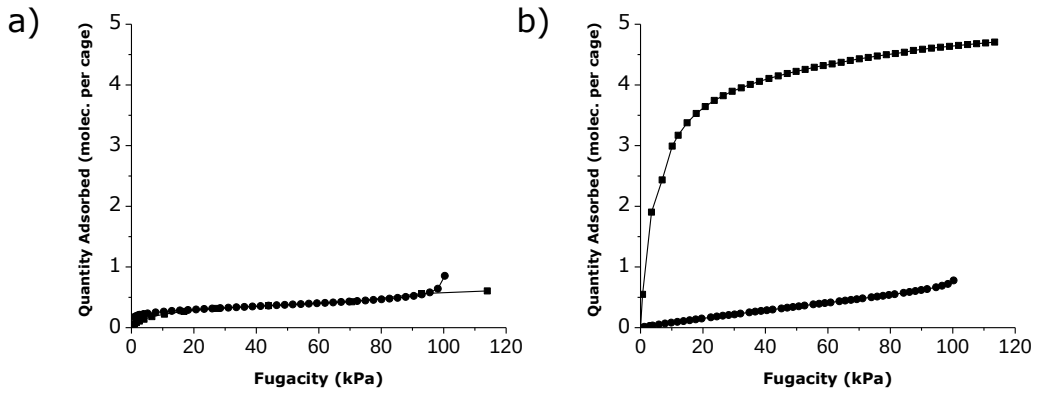


Figure 1.6 – Hydrogen (square) and nitrogen (circles) sorption of **CC1α** and **CC1β**.[103]

Another polymorph of **CC1**, **CC1γ**, was crystallised in the presence of *o*-xylene and is porous to both hydrogen and nitrogen.[103] However, it was not possible to obtain single crystal data for **CC1γ**, therefore X-ray powder diffraction was used to obtain the space group and lattice parameters of the system. This was then assumed to pack in a similar manner to **CC2** (Figure 1.5c), since the space group was also found to be *P*3, and the lattice parameters similar. The crystal motif of **CC2** exhibited the window-to-arene packing that was also observed for **CC1α** (Figure 1.5d), however the methyl groups on the ethylene vertices frustrated the packing of the cages, thus creating an inter-stack one-dimensional channel. This channel has a maximum diameter of 14.0 Å and bottle-necks to a minimum diameter of 6.2 Å.[17]

The observed polymorphism of **CC1** showed that by altering the synthetic reaction conditions, the crystal packing of the cages can be modified, thus having an impact on the properties observed. Chapter 4 focuses on how computational simulations can be used to help understand and evaluate what makes the 'on'/'off' porosity switching of **CC1** possible.

The packing motif of a pair of **CC3** molecules explains how the bulky cyclohexyl groups direct the discrete units into a window-to-window configuration, thus

creating an interconnected diamondoid-pore network (Figure 1.7). When this is combined with the tetrahedral cage symmetry, an interconnected three-dimensional diamondoid pore topology is observed, running from cage to cage (Figure 1.7b). These channels are limited by the diameter of the four triangular windows of each discrete cage (3.62 Å). Any diffusion of gases through this system would have to be as a result of cage-to-cage hopping, *via* an interstitial site located between two neighbouring cages. This means that the porosity is exclusively intrinsic.

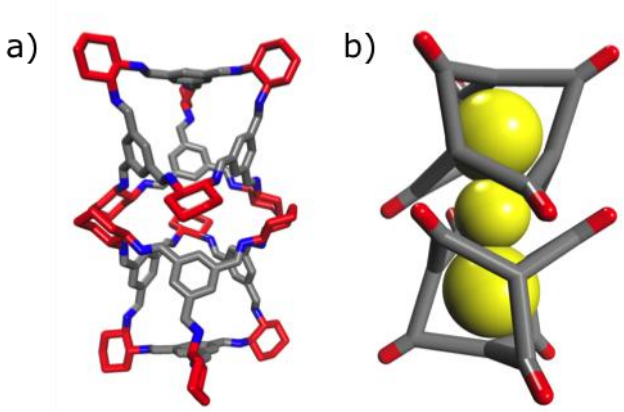


Figure 1.7 – Scheme showing, a) **CC3** dimer pair, with the bulky cyclohexyl groups highlighted in red, directing a window-to-window packing configuration; hydrogens have been omitted for clarity. b) The pore channel (yellow) that exists between the cage dimer, with the cages shown as simplified schematic representations, shown in Figure 1.5.

The crystal structure of each cage was subsequently determined using single crystal X-ray diffraction. Single crystal data for the cages was collected on a Bruker Apex diffractometer, and 1.5 kW graphite monochromated Mo radiation (λ = 0.71073 Å) using 0.3° ω scan steps, spanning at least a hemisphere of reciprocal space for all structures. The frames were subsequently integrated with the SAINT v6.45a (Bruker, 2005).[17] **CC1α**, **CC1β**, and **CC2** were collected at 173 K, while **CC3** was collected at 400 K, so that all of the diffuse solvent contribution was removed. The crystallographic data can be obtained from the Cambridge Crystallographic Database, CCDC, with reference numbers CCDC 707056 (**CC1α**, ethyl acetate solvate), CCDC 720848 (**CC1α**, desolvated), CCDC 782891 (**CC1β**, 2.5 dichloromethane solvent), CCDC783008 (**CC1β**, desolvated), CCDC 720849 (**CC2**, desolvated), CCDC 720851 (**CC3**, chloroform solvate) and CCDC 720850 (**CC3**, desolvated, data collected at 400 K). These are available free of charge at www.ccdc.cam.ac.uk/data_request/cif. More information on the crystallographic data for these structures can be found in previous work.[17,104]

Experimental measurements show that **CC3** exhibited uptake of hydrogen (Figure 1.8a), nitrogen (Figure 1.8b), carbon dioxide (Figure 1.8c), methane (Figure

1.8d), krypton (Figure 1.8e) and xenon (Figure 1.8f).[17] Chapter 5 focuses on modelling the diffusion of gases through **CC3** using molecular dynamics (MD), to see whether these experimental measurements could first be rationalised, and subsequently explained, with a view into highlighting any potential separation capabilities of **CC3**.

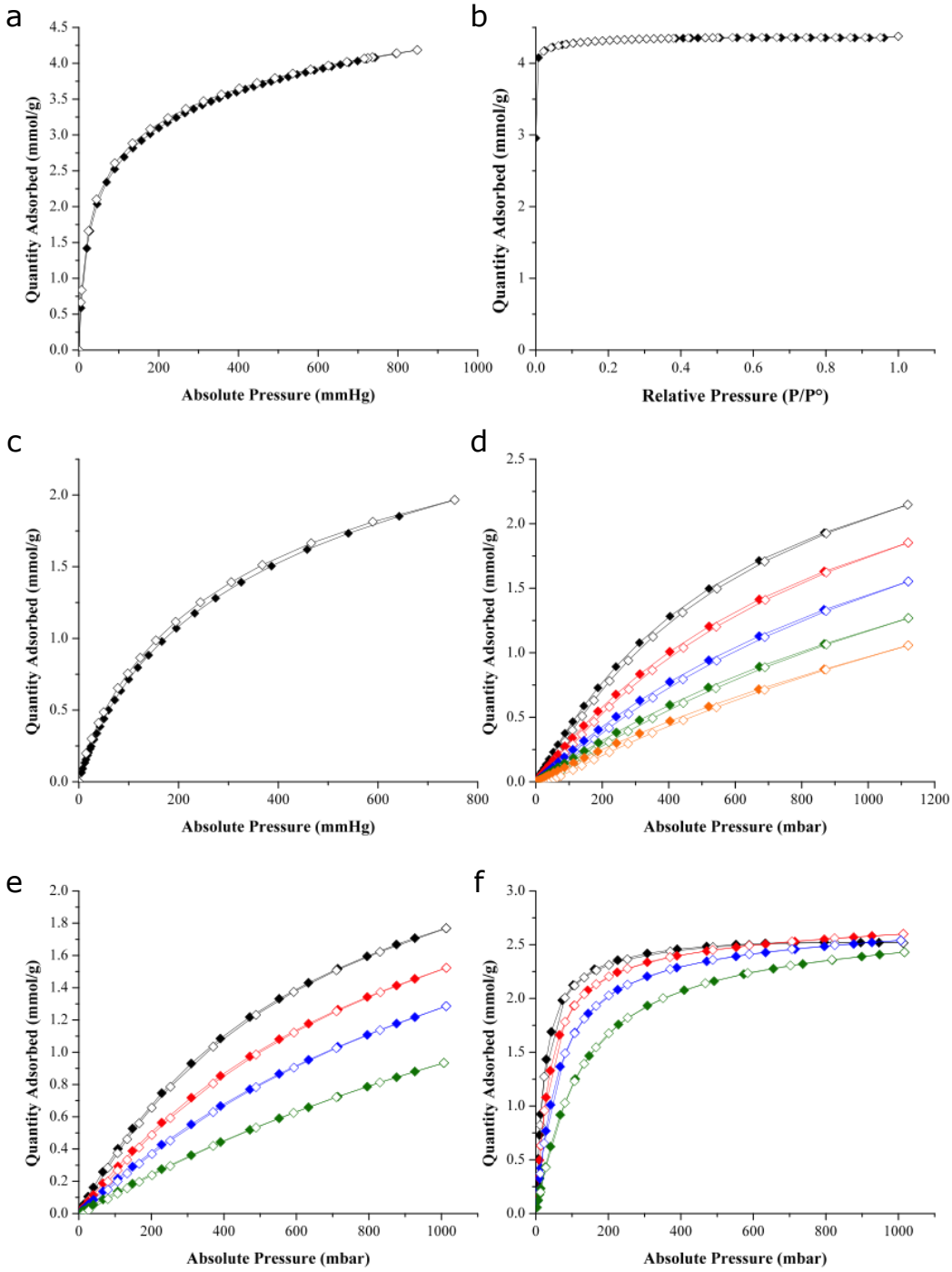


Figure 1.8 – Reversible sorption isotherms for a) hydrogen, b) nitrogen, c) carbon dioxide, d) methane at 263 K (black), 273 K (red), 283 K (blue), 293 K (green) and 303 K (orange), e) krypton at 263 K (black), 273 K (red), 283 K (blue) and 298 K (green), and f) xenon at 263 K (black), 273 K (red), 283 K (blue) and 298 K (green).

Since the discovery of the first three cage molecules, a large family of porous organic cages has now been reported. These include cages with reduced stoichiometry, for example a [2+3] propeller cage, **CC6** (Figure 1.9a),[105] and cages with altered functionality on their vertices, for example the cycloimination reaction of 1,3,5-triformylbenzene with 1,2-cyclopentanediamine, **CC4**.[106] The aldehyde can also be altered to generate larger [4+6] cages, **CC5** (Figure 1.9c),[6] and the stoichiometry can be further increased to synthesise much larger [8+12] cages, for example **CC7** (Figure 1.9d).[107]

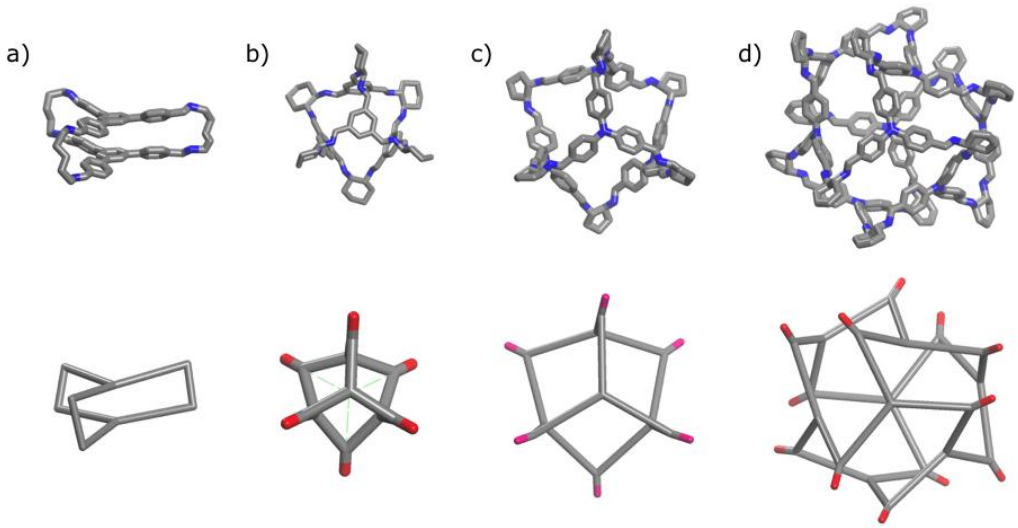


Figure 1.9 – Example of different imine cages synthesised in the Cooper group, a) a [2+3] propeller cage, **CC6**, b) a small [4+6] cage, **CC3**, c) a larger [4+6] cage, **CC5**, and d) a much larger [8+12] cage, **CC7**. Hydrogens have been omitted for clarity. Below each cage is a simplified schematic representation to show the shape of the molecule.

The [2+3] propeller cage, **CC6**, was synthesised *via* the cycloimination of 1,3,5-tri(4-formylphenyl)benzene with 1,5-pentane-diamine. Upon crystallisation, one-dimensional channels were formed, and these showed selective adsorption of hydrogen and carbon dioxide over nitrogen.[105] Selectivity of hydrogen over nitrogen reached 2.4 (1 bar, 77 K), while carbon dioxide selectivity over nitrogen of 11 was reported (1 bar, 300 K). However, the overall uptakes of hydrogen and carbon dioxide were relatively modest, 3.76 mmol g^{-1} and 0.9 mmol g^{-1} respectively, as the one-dimensional channel exhibited a modest SA_{BET} (99 m^2 g^{-1}).[105]

CC5 was synthesised by the cycloimination reaction of tri(4-formylphenyl)amine and (R,R)-1,2-cyclopentanediamine, and is much larger than other [4+6] cages described previously (**CC1**-**CC3**). The cavity inside **CC5** was calculated to be 3.8 times larger than **CC1** and this helps explain a measured SA_{BET} of 1333 m^2 g^{-1}.[6] A

model of **CC5** is shown in Figure 1.9c. It was also thought that the addition of bulky functional groups on the diamine vertices may promote inefficient packing and this may subsequently increase both intrinsic and extrinsic porosity. It has also been shown that some of these cages (**CC1**, **CC2**, **CC4**) can be synthetically interlocked, forming mechanical bonds. Although these lack microporosity, Hasell *et al.* suggested that these may be useful for modular construction of complex porous materials.[108]

The co-crystallisation of three different cages (**CC1**, **CC3**, **CC4**) was accomplished in 2012. The interconversion of **CC1** meant that equal amounts of the *R* and *S* enatiomers exist in solution.[109] When **CC1** was co-crystallised with **CC3**-*R*, a heterochiral co-crystal was generated, exhibiting a SA_{BET} of 437 m^2 g^{-1}. Finally, it has also been demonstrated that it is possible to co-crystallise two enantiomers of the same molecule; **CC3**-*R* and **CC3**-*S* strongly assemble into a co-crystal, and this exhibits a SA_{BET} of 873 m^2 g^{-1}.[24] The racemic **CC4**-S/**CC3**-R co-crystal was synthesised, and this enhanced the surface area further, increasing the SA_{BET} to 980 m^2 g^{-1}.[6]

1.7. Comparison of porous networks and porous molecules

One advantage of porous networks is that they are assembled using strong covalent bonds, and these improve the stability, also increasing the pore stability upon evacuation.[110] These networks tend to have low densities and consequently high surface areas, with SA_{BET} reaching over 5500 m^2 g^{-1}.[37,111,112] In addition, it is possible to tailor the pore diameter within these porous networks, making it possible to target specific gas separations.[50]

Although porous molecules to date do not exhibit as high surface areas as porous networks, they do have many potential advantages over network materials.[22] Four examples are:

1. **Solubility**; porous molecules can typically be dissolved in common organic solvents. This enhances their processability, allowing their structure to be modified *via* recrystallisation, switching properties on and off.[104] In addition, these can be cast as porous films.[22,63]
2. **Synthetic diversity**; it is possible to modify the functional groups on discrete molecules when compared to heterogeneous transformations in porous networks.[22,100,113]
3. **Mobility**; the majority of porous molecules lack the strong covalent or coordination bonds that hold porous networks together. This means that

there is scope for porous molecules to respond to the sorption of guests in ways not accessible to porous networks. This could include cooperative encapsulation of guests within the molecule itself, or the extrinsic void volume, and has potential applications in carbon capture, molecular recognition or detection and the removal of atmospheric pollutants.[22]

4. **Mix-and-match strategies**; due to their soluble nature, it is possible to mix a combination of different porous molecules together to exist in one system. This could achieved by co-crystallisation of multiple discrete subunits.[114] This could greatly affect the pore structure and properties of the single entities on their own, leading to a wide range of applications, including the potential synthesis of a material containing both isolated acid and base groups by encapsulating both groups in separate molecules and subsequently co-crystallising both subunits. This variation of functionality would mimic biological structures, for example enzymes.[22]

Overall there are many potential advantages in the use of porous molecules over porous networks, and finding and exploiting these is a significant focus for research in porous molecular materials.

1.8. Applications of porous materials

One factor attributed to the rise in global warming is the overuse of fossil fuels, which has resulted in a high production of carbon dioxide.[115] This intensive use of fossil fuels has led an increase of over 30% in the concentration of carbon dioxide in the earth's atmosphere.[116] This increase in concentration has been predicted to increase substantially over the next two decades, therefore making it imperative to devise ways to capture, or store, the excess carbon dioxide to protect the environment.[117] Consequently, a vast amount of time and research has been spent on solutions to tackle this problem. One method to reduce the levels of carbon dioxide in the atmosphere is to selectively separate or capture carbon dioxide from the air. This can be achieved in two ways:

- **Pre-combustion:** the removal of carbon dioxide from fossil fuels before they undergo combustion, for example, in a power plant.[20]
- **Post-combustion:** the removal of carbon dioxide after fossil fuel combustion, for example, from exhaust gas.[20]

Alternately, it has been shown that one of the most efficient ways to reduce carbon dioxide emissions is to develop revolutionary changes in the technology of energy production, distribution, storage, and conversion.[115] Many techniques

have been suggested, and one possible way is to use an alternative energy to fossil fuels, for example hydrogen.

At present, hydrogen is mainly produced through the reformation of steam-methane, followed by a water-gas shift reaction.[116] The composition of this is around 71% hydrogen, 15% carbon dioxide, 4% methane and 1% carbon monoxide, with the rest being made up of other gases, such as water.[118] It is important to devise techniques to separate hydrogen and carbon dioxide from the effluent gas.

To date, the majority of hydrogen is stored using two methods - liquid storage at low temperatures and gas storage at high pressures. There are however, major drawbacks to these approaches, such as high expense, and concerns in the safety of storing hydrogen at high pressure.[119] An alternative is to use porous materials, such as zeolites, to both store and separate hydrogen and carbon dioxide.[119] Zeolites are highly stable, low cost, and exhibit a wide range of pore structures; this had led to research into their ability to both store and separate hydrogen and nitrogen. Two main methods of storing hydrogen in zeolites are encapsulation and adsorption:

- **Encapsulation**: Also know as trapping, this is achieved by forcing the hydrogen molecule to occupy voids within the pore structure that are not usually accessible under normal conditions, by increasing the temperature and/or the pressure of the system.[119] When the system is returned to normal conditions, the gas is unable to be released. To subsequently release the gases from the porous host, the temperature is increased.[32,120,121]
- **Adsorption:** This is the adhesion of gases to a solid surface. For hydrogen, cryogenic temperatures are needed to achieve reasonable loadings, as the interaction energy between hydrogen and the solid surface is very low (less than 10 kJ mol^{-1}).[119] Experiments by Vitillo *et al.* have shown that hydrogen capacity in zeolites is intrinsically limited to around 2.86 wt%, and this is due to geometrical constraints – this is well below the DoE targets.[119]

For hydrogen to become a *via*ble alternative to fossil fuels, a safe, high-capacity system would be required, however no conventional method is commercially available. The DoE have set a target for hydrogen storage of 45 mmol g^{-1} (9 wt%),[122] and this target is still to be met by porous materials.

One of the most studied MOFs is MOF-5, which was originally reported by Yaghi *et al.* in 2003, and is shown in Figure 1.2b. The reaction of zinc clusters with 1,4 benzenedicarboxylate (BDC) generated a three-dimensional cubic extended porous structure.[41] This displayed a high uptake of hydrogen, 22.5 mmol g^{-1} (1 bar, 78 K), equivalent to 4.5 wt%.[41] Large amounts of hydrogen storage have also been observed in other MOFs, with Furukawa *et al.* also demonstrating uptakes of up to 37.5 mmol g^{-1} in MOF-177 (70 bar, 77 K), equivalent to 7.5 wt%.[123] Yan *et al.* then exceeded this by showing hydrogen uptake of 50 mmol g^{-1} in NOTT-112 (77 bar, 77 K) and this corresponds to 10 wt%.[124]

Another alternative to fossil fuels is methane, and this is desirable as it burns more cleanly than gasoline, and possesses the highest hydrogen to carbon ratio, when compared to other hydrocarbons.[125,126] In 2010, around 23.8% of the global energy consumption was supplied by natural gas, and if this trend continues, the volume of natural gas consumed will increase, especially with the imposition of financial penalties on carbon emissions.[127] This could reach up to 50% over the next twenty years.[127] The composition of natural gas is predominantly methane, although other heavier hydrocarbons are present, as well as nitrogen. One problem with natural gas is that a large amount of the reserves contain a high level of nitrogen and carbon dioxide.[128] These levels must be reduced to less than 3% nitrogen and 2% carbon dioxide if the gas is to be used in pipelines.[129] Therefore it is important to find ways of reducing the high levels of nitrogen and carbon dioxide. One way to accomplish this is to use porous materials to separate methane from carbon dioxide and nitrogen.[127]

The major obstacle with using methane as a fuel, is the amount of storage space required, as the compressed gas requires three times as much space when compared to an equal volume of gasoline. This means that the discovery of materials that enhance methane storage could aid the use of natural gas as an alternative to fossil fuels. The DoE set the target of storing 180 STP (standard temperature and pressure) litres of methane per litre of storage, and this has already been exceeded by a number of MOFs.[125,130,131] This corresponds to about 35 wt%, and is necessary to compete with compressed natural gas.[132] One advantage of methane over hydrogen, is that a fair amount can be absorbed at room temperature.[132] Eddaoudi *et al.* showed that IRMOF-6 exhibited a methane uptake of 10.7 mmol g^{-1} (36 atm, 298 K),[40] while Millward *et al.* reported that MOF-177 exhibited a carbon dioxide uptake of 33.5 mmol g^{-1} (42 bar, room temperature).[133] Llewellyn *et al.* surpassed this in 2008 when they reported a carbon dioxide uptake of 40 mmol g^{-1} in MIL-101 (5 MPa and 303 K). One reason

for high uptake of gases in MOFs is their potential for high surface areas and void volumes; Furukawa *et al.* reported SA_{BET} and SA_{LANG} of 6240 and 10400 m^2 g^{-1} respectively for MOF-201.[112]

Recently, noble gases, especially krypton and xenon, have attracted a large amount of interest, due to their use in industry, as a result of their physiochemical properties.[134] The pure form of both krypton and xenon are valuable commodities that have a wide range of uses, for example medical imaging, anaesthesia and lighting.[135-137] The separation of krypton from xenon is also an industrially relevant problem.[138] Both krypton and xenon are used in fluorescent lights, and currently these gases are isolated using cryogenic distillation. As there is only a small amount of these present in air, 1.14 ppm krypton, and 0.086 ppm xenon, this is an energy intensive process. Both gases are separated off into the oxygen-rich stream post distillation, and are subsequently purified to obtain a 80:20 molar mixture of krypton to xenon.[139] This then undergoes further cryogenic distillation to produce the gases in their pure form, thus materials that selectively adsorb krypton or xenon are highly desirable, as this would reduce cost.[138] The separation of krypton from xenon is also an important step in removing radioactive ^{85}Kr from spent nuclear fuels, and this also includes a cryogenic step, therefore making it highly energy intensive.[45,139] However, even after cryogenic distillation, the radioactive traces of ^{85}Kr are still too high for the xenon-rich phase to be used commercially.[45] If the concentration of ^{85}Kr could be reduced to acceptable levels, a new source of xenon could become available for commercial use. One potential method is the use of porous materials that selectively absorb one noble gas over the other.

In 2012, Thallapally *et al.* tested two MOFs, MOF-5 and NiDOBDC, for xenon capture and separation. Results highlighted a high uptake of xenon in NiDOBDC, as well as a high selectivity of xenon over krypton, beating that observed in activated carbon.[140] Thallapally *et al.* also showed promising results for the removal of xenon and krypton from nuclear fuel reprocessing plants, using Ni/DOBDC – a Kr/Xe selectivity of 7.3 was reported.[45] In addition, the Kr/Xe selectivity in two partially fluorinated MOFs, FMOFCu and FMOFZn, has been reported to be influenced *via* a temperature change. FMOFCu exhibited an inversion in selectivity toward krypton at temperatures below 0 °C, though this was not observed for FMOFZn.[141] Research into noble gas selectivity has not been limited solely to MOFs, as investigations into the sorption and separation of noble gases in zeolites, and activated charcoal has also been reported.[142-144]

1.9. Modelling porous materials

It is possible to use computational techniques to help understand and rationalise experimental results, as well as suggest potential synthetic targets *a priori,* using structure prediction methods. For porous materials, properties such as surface area, gas sorption isotherms, uptake and separation can be simulated. In addition, there is potential to screen both synthetically available and hypothetical structures for advantageous properties using high throughput techniques.

1.9.1. Describing the energetics of porous materials

There are a number of approaches available for describing the energetics of porous materials. These include coarse-graining methods,[145] electronic structure calculations and classical force field (FF) methods. Course graining tends to be used on large, often amorphous systems,[146] and is not the focus of this work. Electronic structure calculations use quantum mechanics to represent the electronic distribution of a system, thus giving information about optical, electronic and magnetic properties.[25] The computational expense of high-level wavefunction methods is large, and this limits the size of the system that can be studied with such methods; one way to reduce this computational expense is to use density functional theory (DFT). Electronic structure calculations are discussed in more detail in Chapter 2.

A great deal of work has gone into the development of FFs suitable for predicting the structure and dynamics of a system. These ignore electronic degrees of freedom and refer to a set of potential parameters and functional forms used to describe the potential energy surface (PES) of a system; these are often derived from a combination of fitting to experimentally determined properties, and quantities determined through high-level quantum mechanical calculations. The advantage of such FFs is that they are computationally less expensive than other approaches. The disadvantage is that the accuracy can suffer if they are not parameterised correctly. Therefore, there has to be a balance between accuracy and efficiency.

There are a wide range of 'off-the-shelf' FFs available, and these aim to be as transferable as possible. These are only appropriate for qualitative analysis and not an in depth quantitative analysis,[147] which requires the FF to be explicitly fitted for the both the desired system, as well as the purpose of the simulation. Examples of 'off-the-shelf' FFs include the Universal force field (UFF),[148]

COMPASS,[149] and the polymer consistent force field (PCFF).[150] Descriptions of these are found below:

- **UFF:** The aim was to facilitate the simulation of a wide range of atomic structures. Therefore UFF is based on a set of general rules for estimating the FF parameters, and these are determined by the element present, its hybridisation, as well as its local connectivity.[148] Initially, calculations were parameterised using a few organic molecules, and once these were fitted, three main group inorganic molecules were parameterised:
    - 1,3,5,7-tetrakis(trifuoromethyl)-2,4,6,8,9,lO-hexathia-l,3,5,7-tetragermaadamantane,
    - Octamethylcyclotetrasiloxane,
    - Dodecaphenylcyclohexastannane.

  Subsequently, a wide variety of transition metal complexes were examined.
- **COMPASS:** This was designed for the atomic simulation of common organic molecules, inorganic small molecules, and polymers.[149,151]
- **PCFF:** This has been applied to a wide range of systems including polycarbonates, melamine resins, polysaccharides, other polymers, organic and inorganic materials, as well as carbohydrates, lipids, and nucleic acids. Based upon CFF91, PCFF introduced a number of additional parameters by supplementary fitting of polymers by Sun *et al.*,[150,152,153] as well as fitting of zeolite framework fragments by Hill and Sauer, who used DFT calculations of small, representative zeolite fragments, to generate reference data.[154]

As well as generic 'off-the-shelf' FFs, which could be used for porous materials, there are several examples of FFs that have been accurately parameterised for specific porous materials. Schmid *et al.*[155] recognised that adsorption and diffusion studies for COFs tended to use generalised FFs where the structure was constrained to that seen experimentally. Consequently, they developed a FF to address this, using a genetic algorithm-based parameterisation approach, which was the first fully flexible model for COFs.[156] Much work has also focused on IRMOF-1, where the flexibility of the framework was taken into account by adapting the CVFF basic potential functions based on *ab initio* calculation;[157] Greathouse *et al.*[158] then used this FF to investigate IRMOF-1's thermal conductivity. More recently, FFs have been developed for zeolitic imidazolate frameworks (ZIFs) so that their flexibility and gas adsorption[159] can be studied in greater detail. In this thesis, a cage specific FF (CSFF) has been developed, this is discussed in depth in Chapter 3.[160]

1.9.2. Predicting the structure of porous materials

Although the discovery of porous materials has traditionally been serendipitous, there have been efforts to guide their synthesis. Mastalerz *et al.* reported one such example for porous molecular materials in 2012, when they rationalised extrinsic porosity *a priori* by selecting molecules with directional bonding; upon crystallisation a high surface area was observed.[90] When considering porous molecules, it is possible, in principle, to use polymorph prediction techniques; these were originally developed for pharmaceutical molecules,[161] to predict the energetically favoured crystal motifs.[25] If *de novo* predictions become reliable, these could be used to assist determination of structures where it proves difficult to obtain single crystal diffraction data. The major difference between pharmaceutical molecules and porous molecules is their size. As porous molecules are typically much larger (e.g. > 100 atoms for porous imine cages of Cooper group), these calculations are much more difficult to undertake, and can often take weeks.

Contrary to this, there are a few ways to predict the structure of porous networks, for example the enumeration of topologies using an isoreticular structure.[162-164] In addition, structures can be generated using building blocks.[165,166] The energetic ranking of porous materials in structure prediction also differs, as there are reliable, computationally inexpensive methods to rank hypothetical porous networks, for example zeolites,[167] by their relative stabilities. However, the presence of metals can complicate these calculations, especially for MOFs, as multiple valences make it difficult to choose an appropriate FF, resulting in limitation to higher level, computationally more expensive methods; this makes it impractical to calculate all the predicted structures.[165] The disparity between the approaches required for structure prediction of porous networks and porous molecules is down to a number of factors. For example, non-covalent or non-coordinative intermolecular forces dictate the packing of molecular crystals, as they are discrete in nature.

In comparison, the energetics for the polymorphism of porous molecules is largely dictated by contributions from the packing forces between the discrete units and any intramolecular strain required to hold the molecules in their specific crystal motif. Consequently, reliable FFs are either available, or can easily be refitted, to measure such energies.[25] To run CSP on a molecule, a number of key steps are implemented, and these are outlined below:

1) The structure of the molecule is optimised, using electronic structure calculations, so that the starting geometry and partial charges of the molecule can be ascertained.[168]
2) The optimised molecule then undergoes Monte Carlo simulated annealing. To save computational expense it is kept rigid, and only the most common space groups are applied.
3) The simulated structures are then clustered to remove duplicates and ranked in terms of lattice energy.
4) A subset of lowest energy structures are then optimised using higher level energy calculations. In some cases, these are allowed to relax, as this is important to obtain the energetics of flexible molecules.
5) Finally, these are re-ranked, using the new lattice energies.

It is possible to use solid state DFT calculations for the whole process of CSP, though this is computationally expensive, thus limiting it to only small molecules (less than 100 atoms).[169] Assuming that no templating or solvation effects are observed experimentally, the predicted low energy structures should contain any experimentally realised.

To date, there are limited examples of successful crystal structure predictions for porous molecular materials reported in the literature. In 2009, Day *et al.* showed successful predictions of four extrinsically porous molecules; Dianin's compound, urea, bicyclodiol and hydroquinone.[170] Although this was a new direction for the use of CSP, the molecules were still relatively small in size. However, in 2011, CSP was successfully used to predict the packing of much larger imine cages, and at present these are the largest known successful CSPs.[6] Calculations showed that the crystal structure of two **CC1** polymorphs, **CC1α** and **CC1β**, had similar lattice energies, and this would explain why **CC1** exists in multiple polymeric forms. The same work also successfully reproduced the **CC3** crystal structure while correctly indicating that there was a strong preference for heterochiral packing; the energy difference between the heterochiral and homochiral packing was 32.1 kJ mol^{-1}.[6] This shows that there is definitely potential in the CSP of porous molecules and this could, in principle, be used as a guide for synthesis, as well as analyse potential packing motifs for desirable properties before attempting to synthesise them in the laboratory.

It is also possible to predict the structure of amorphous porous materials, though much less work has been done in this area. Amorphisation of porous materials can be achieved by design, for example using rapid freeze-drying techniques,[24] or

when the crystallisation of porous molecules is prohibited by the shape and size of the discrete units. Porous carbons were one of the first amorphous materials to be modelled, while recent attempts to predict the theoretical limit of surface area for molecular materials have been undertaken using carbon disc clusters.[171]

Very few computational predictions on porous amorphous solids have been reported, however Colina *et al.* have shown that it has been possible to predict the packing of PIM-1 *via* computational screening.[172] Here a 21-step MD compression and relaxation technique was used, with the pore size distributions, surface areas, structure factors, and sorption isotherms subsequently characterised for simulated structures; these all showed good agreement to experimental results. Cooper *et al.* used a similar approach to simulate the structure of amorphous 'scrambled cage' cage mixtures, where a cell was loaded with an experimentally determined ratio of cage species, using a Monte Carlo sampling technique, and subsequently energy minimisation until the experimental, target density was achieved.[26] The major limitation with this work was that, at present, the simulations needed experimental input, and this makes *a priori* predictions difficult. However, these methods are likely to improve over time, so that no experimental properties are required, thus allowing true predictive techniques to, in principle, be implemented on amorphous molecules; this has already been achieved with amorphous networks.[173] In fact, Colina *et al.* have just reported an example of this.[174]

1.9.3. Modelling guests in porous materials

Computational modelling can be used to identify new targets for gas adsorption, separation and storage. If characterised, the single crystal X-ray diffraction structure can be used to represent the material, however if no experimental data is available, it is, in theory, possible to use structures generated *via* CSP.[6] Simulations of gas docking are typically calculated using Monte Carlo techniques, where the sorption sites within the material are randomly sampled using a FF. This allows both the sorption uptake and adsorption enthalpy to by calculated over a range of pressures.[25] The same methods are used for porous networks and porous molecules. [125,175] It is also possible to use Grand Canonical Monte Carlo (GCMC) to produce full adsorption isotherms and this has been widely used for *in silico* screening of porous networks, especially when attempting to identify materials for selectivity purposes; this is covered in more detail in Section 1.9.4.[176] GCMC simulations allow the determination of preferred sorption sites, as well as inferring which parts of the pore topology contribute towards the porosity

of the system. It can also be used to calculate guest selectivity and competitive sorption.[25] It is important to consider the following when comparing simulated adsorptions to experimental uptakes:[177]

1) One cannot assume that good agreement over a narrow pressure range is sufficient, as while there may be good agreement at a low pressure range, simulations and experimental data can vary to a larger extent at higher pressures.[125]
2) The quality of the experimental data and the accuracy of the structure used can have an impact on the properties and uptakes reported. For example, full evacuation of solvent from the material can lead to imperfections in the structure.[125]

At present, Monte Carlo (MC) simulations are typically carried out using rigid hosts, though this proves more difficult when modelling porous molecules, due to the increased mobility of their discrete molecular units. In addition, this makes it harder to model stepped isotherms, as the host structure often changes in response to the guest. This has been addressed in MOFs by modelling the uptake of both the empty and fully loaded crystal structures; the results have subsequently been used to explain the experimental stepped isotherm.[178] Single component isotherms have also been calculated for carbon dioxide and methane in TBC4,[179] and these suggest that structural changes or defects within the experimental structure may be partially responsible in producing relatively higher simulated adsorption, when compared to experiential measurements.[179,180] However, this approach is reliant on the crystal structures of both phases being available, and this is not always the case. This has led to theoretical studies on phenomena such as gate-opening transitions,[181] breathing of the host material,[182] and swelling of the host structure to solvent adsorption.[183] One issue with GCMC simulations is that time-dependent effects are not included, and this has been shown to have a large effect on the properties of porous materials, for example, increased pressure can influence the pore size.[184]

Therefore, effort has been made to include time-dependent effects into the calculations, and this has been achieved by adding in a MD step after each guest is loaded. These hybrid MC/MD simulations aim to model previously impossible phenomena, such as the guest-assisted structural transitions seen in the porous material that correspond to large breathing effects in the structure. Another limitation of sorption calculations is that no consideration is taken into the connectivity of the pore topology. GCMC arbitrarily docks guest molecules into

the host system, leading to possible guest inclusion in sites where diffusible pathways are not permitted.[185] Even after taking these limitations into account, simulated guest sorption is still a very powerful tool, and the literature is full of examples where computational studies have aided the rationalisation of experimental findings, including adsorption, selectivity, and separation.

Due to methanes potential as an alternative to fossil fuels, the adsorption of methane in porous materials has attracted a large amount of interest. Düren *et al.* carried one of the first studies that compared a number of porous structures, and this focused on the adsorption of methane into 18 different materials, including MOFs, zeolites, molecular squares, MCM-41 and carbon nanotubes.[185] The aim was to rationalise the interplay of the factors influencing methane adsorption: surface area, free volume, strength of host/guest interaction energy, as well as pore size distribution.[185] They concluded that to achieve high methane storage, a combination of factors were desirable, including a large surface area with a high free volume, low framework density, and strong methane-adsorbent interaction.[185] Wang also reported a systematic study on methane absorption in ten MOFs, all with different topologies.[186] GCMC simulations were used (up to 100 bar, 298 K) to determine the ideal characteristics for optimal adsorption of methane. Wang concluded that there was a strong correlation over the high pressure range between methane uptake, high surface area, and free volume.[186]

Gallo and Glossman-Mitnik also calculated adsorption isotherms for methane in a number of MOFs. Here two large surface area MOFs, two catenated MOFs, and a MOF with open metal sites were compared (up to 80 bar, 298 K).[187] They found that at low pressures, the largest gravimetric uptake of methane was seen in the catenated MOFs, whereas for higher pressures the large surface area MOFs showed the highest uptake; this was subsequently linked to the higher methane adsorption enthalpy.[187] Several methane adsorption studies have been reported for COFs.[4,51]

Selectivity in porous materials using computational techniques has been reported for carbon dioxide and carbon monoxide in ZIF-68 and ZIF-69. GCMC calculations were used to simulate the adsorption isotherms for the two ZIFs (up to 1 atm, 273 K).[188] The simulations showed that carbon dioxide associated strongly with the benzene rings from the benzimidazolate anion, however the binding energies of carbon monoxide were much smaller. The binding sites were also calculated, and the nature and number of these sites in the channels and pores of the ZIF structures were determined from the guest probability distributions. The

adsorption isotherms showed good agreement with experimentation for carbon dioxide, though overestimated the carbon monoxide adsorption isotherms.[188]

Interest in mesoporous adsorbents, like MCM-41, has also increased, especially in the separation stage of carbon capture and storage.[180] Research into the effectiveness of functionalising the surface groups to tune CO_2/N_2 separation, for use in a power plant flue gas stream, has been carried out by Williams *et al*. The adsorption isotherms of the gases in MCM-41 were calculated using GCMC simulations. The carbon dioxide capacities and selectivities were subsequently evaluated, so that their applicability as potential materials for the removal of carbon dioxide from flue gas could be assessed;[180] this was achieved by comparing their pore size, surface group concentration, surface group rigidity and alkyl chain length. This highlighted that the addition of tailor-made functional groups to MCM-41 led to increased carbon dioxide selectivity, therefore these materials should be targeted for carbon-capture applications.[180] The nitrogen adsorption for MCM-41 was also calculated by Coasne *et al*.[189]

It is also possible to simulate the diffusion of guests through porous solids, using MD simulations. The energetics of the systems can be modelled using both electronic structure calculations and FF techniques. *Ab initio* MD simulations might, in theory, be more accurate than the molecular mechanics approach, though *ab initio* calculations are computationally expensive, and this can be prohibitive when wanting to specific calculate properties that require long time periods, for example diffusion coefficients. This work therefore focuses on FF based MD simulations. These have been used to simulate the behaviour of porous materials, so that their properties could be analysed; phenomena like gas sorption, gate opening, and diffusion can be examined.[190]

As the majority of microporous solids are extended networks (such as zeolites, activated carbons, MOFs,[4,9,11] COFs,[4] or polymer networks), when simulating sorbate diffusion, the host structure is often considered to be rigid. This 'rigid host assumption' allows relatively fast and computationally inexpensive calculations, but discounts any flexibility in the host system. For zeolites, this broadly holds so long as the radius of the sorbate is significantly smaller than the smallest pore aperture.[4,9,11] This methodology, though, breaks down when the pore aperture becomes similar in size to the guest.[2,4,19,71,191] Hence, the assumption of rigidity is best restricted to materials with low densities and large pores, such as certain mesoporous MOFs or COFs. However, there are many microporous solids in which the flexibility of the host can have a profound effect

on the sorption behaviour. A well-known example is the MOF, MIL-53,[192] which displays profound structural changes in response to guest adsorption. This can expand or shrink in response to adsorption of guest molecules, for example carbon dioxide and water, therefore the inclusion of flexibility for the host material would be vital if MD simulations were to be calculated.

Modelling the dynamic nature of porous materials allows the influence, if any, that an inserted guest has on the hosts pore structure to be analysed. In addition, MD simulations allow the pore necks in the system to be monitored, with transient channel openings examined, and this could explain why a system shows experimental uptake of a guest, while being conventionally non-porous.[21] The introduction of flexibility into the simulations allows the pore topology of the system to evolve over time, and this could lead to systems exhibiting "porosity without pores". The breathing modes and vibrational frequencies of the host material can also be simulated and used to validate experimental measurements. Sholl *et al.* have highlighted that it is possible to monitor the window size of ZIF-8 as a function of frequency for a short *ab initio* MD simulation; this showed how the fluctuation of the window evolved during the simulation and aimed to describe how this flexibility affected the diffusion of both methane and carbon dioxide.[193]

In 2011, Zhao *et al.* reported that although there is an extensive amount of work already in the literature describing the adsorption and diffusion of gases through MOFs, most of these were fixed in their experimentally determined crystallographic positions.[194] Normally 'off-the-shelf' FFs were used in combination to the rigid hosts, but Zhao points out that since MOFs may exhibit substantial changes in unit cell volume upon external stimulation by a guest or outside factor like temperature, FFs that describe the dynamic nature of MOFs are needed, so that the frameworks no longer have to remain constrained.[194]

One issue is that these 'off-the-shelf' FFs may not accurately describe the nature of the porous materials. Avellaneda *et al.* have recently published a paper outlining controlled porosity in robust organic cage materials. The MD simulations for this were performed using UFF, and as this was not designed to describe the dynamic behaviour of the cages, it could lead to inaccuracies in the results and over analysis of the data.[195] Xue *et al.* acknowledged that frame flexibility played a key role in understanding the diffusion of hexane in MOFs and therefore modified the MM3 FF, so that the flexibility of the framework could be accurately taken into consideration. It was clear that the diffusion pathway of hexane was

largely affected by gas loading, in combination with the structures of the MOF pores. For systems where the host-guest interactions were weak, increased loading did not affect framework flexibility, with the effect of temperature was more influential. It was also evident that the flexibility of the MOF framework was determined by the nature of the organic linkers.[196]

A number of comparisons have been made between the fixed framework, and FF described MD simulations. Smit *et al.* compared results for rigid and flexible zeolites, and concluded that although the flexibility had very little effect on the thermodynamic properties, there may be a more pronounced effect on transport properties, as the flexibility may increase or decrease the free energy barrier for diffusion.[4,9,11,197] Combariza *et al.* also reported the importance of including framework flexibility in zeolites, stating that this has a big impact on the diffusional behaviour of adsorbates in nanoporous materials.[198] It increases the potential energy barrier for gas passage through the material.[2-6,199] Nemkevich *et al.* have recently simulated guest transport in clathrates of Dianin's compound and hydroquinone. They explained that as flexibility in rigid structures, like zeolites, is important in allowing guest adsorption and transport to occur, it must be at least as vital in dictating guest diffusion in organic host lattices, as these are only held together by weaker intermolecular forces.[4,9-11,200]

It is also possible to use MD in the gas phase, focusing on the intramolecular motion of one, or a couple of discrete molecules. One such case is an investigation into the behaviour of a single calix[4]arene molecule.[201] Using the AMBER FF, Adams *et al.* studied the behaviour calix[4]arene as different guest molecules were introduced into the cavity, and monitored how well the guest was retained when the complex was allowed to evolve with time.[201] They noticed that when a small molecule formed a stable complex, the guest orientated in such a manner that any positive charge density aligned with the negative charge density of the hosts aryl rings.

To date, both GCMC and MD techniques can be used to give insight into what is observed experimentally, rationalising phenomena such as gas uptake, gate opening, and 'porosity without pores'. In principle, these approaches could be applied to hypothetical materials to predict candidates for synthesis that exhibit advantageous properties. For this to be achievable, databases of hypothetical porous materials are required. Generation of such databases is discussed in the next section.

1.9.4. Hypothetical screening of porous materials

As computational methods for screening porous materials become more powerful, a large amount of research is going into the generation of databases of hypothetical porous materials. These could then be screened for advantageous properties, and one such example is when Krishna *et al.* reported the *in silico* screening of MOFs for separation applications.[176] Here, calculations based on earlier work,[202] outlining the use of configurational-bias Monte Carlo (CBMC) and MD simulations was used to screen a wide variety of MOFs, ZIFs, and zeolites, for separation of a variety of mixtures. One of their conclusions was that from all investigated MOFs, the best carbon dioxide capture was observed in materials that offered strong host-guest interactions between the carbon dioxide molecule and the exposed metal sites.

Another way to identify such materials is to screen libraries of thousands of hypothetical materials for specific criteria. Building upon previous work,[138] Snurr *et al.* screened over 137,000 hypothetical MOFs for selectivity of krypton over xenon.[203] They used multi-component GCMC simulations and calculated the pore limiting diameter, largest cavity diameter, accessible void volume and the adsorption of krypton and xenon at 1.0, 5.0 and 10 bar, 273 K.[203] As they were interested in the thermodynamic separation of krypton from xenon, they used the kinetic radius xenon, 4.1 Å, as a cut off, assuming that materials with a pore limiting diameter (PLD) close to this could, in principal, separate krypton from xenon; for krypton they used a kinetic radius of 3.6 Å. This assumes that krypton would be small enough to diffuse through the hypothetical MOF, whereas xenon would be restricted due to its radius and therefore not able to diffuse through the pore structure of the system.[203] They conclude that by using a combination of geometric and thermodynamic analysis, they were able to confirm a previous hypothesis that cavities that could adsorb a single xenon atom would display the highest selectivity. In addition, they proposed hypothetical structures that exhibited higher selectivities than any existing MOF.

Recently Snurr *et al.* have also used a combination of breakthrough measurements and GCMC simulations to show that MOF-505 is an ideal candidate for separating xenon from krypton.[204] A measured Xe/Kr selectivity of 9-10 was reported and this is among the highest to date. This is rationalised by the pore confinement effect of the small pores running through the system. They also noted that breakthrough measurements showed that HKUST-1 only had a modest Xe/Kr selectivity of 4.5. They concede that although GCMC simulations do provide

useful suggestions for adsorption separations, the kinetic effects are not taken into consideration, and these may play a vital role, therefore breakthrough measurements are a useful method to validate any computational findings.

A large amount of work on hypothetical screening has also been undertaken by Sholl *et al.* Initially, they developed tools to aid in the calculation of diffusion limitations in MOFs, with the aim of identifying materials ideal for kinetic separation.[205] These tools were then applied to a large set of hypothetical zeolites, to provide information on specific properties, for example the largest cavity diameter, as well as the range of adsorbates that could feasibly diffuse through each zeolite. The hypothetical zeolites were then compared experimentally know zeolite crystal structures.[206] The vision was to build an integrated series of modelling tools that could rapidly screen a large set of materials, highlighting those with advantageous properties. Subsequently, more detailed computationally expensive methods could be applied to this small subset of data. Sholl *et al.* then reported the addition of host-guest interactions of non-spherical adsorbates, nitrogen and carbon dioxide into their screening tools, and subsequently screened a large set of MOFs for selectivity. The PLD was also used as a potential screening method.[207]

In addition, Sholl *et al.* were also interested in the separation of rare gases using MOFs. They sampled a collection of over 3000 MOFs to look at their performance in binary separation of rare gases.[208] They combined ideal adsorbed solution theory (IAST) and MD simulations to predict which of these MOFs would be ideal candidates for gas separation. The frameworks were kept rigid during the MD simulations and the PLD was monitored. First they aimed to identify candidates for separation of noble gas pairs, and subsequently determined whether binary IAST calculations were applicable. Finally they sought to rationalise why certain MOFs displayed good separation, with a view to identify additional candidates through further screening. They reported that in almost all cases IAST was able to accurately calculate binary selectivity, and that this only fails due to the occasional blocking of pores by one of the two adsorbing species. They also noted that their materials showed that many of their materials exhibited an unexpected 'reverse' selectivity; that is, adsorption of the smaller species is preferred.[208]

Both Snurr and Sholl found that the PLD was helpful when trying to discern on certain screening criteria, however as they utilised only GCMC simulations, using a static host structure, it is not clear whether the PLD is applicable when utilising MD with a fully flexible system. This is because the PLD will change as the

simulation evolves. It is necessary to assess whether the flexibility of the system is important.

It is also possible to characterise the pore landscape of materials using computational software that automates the analysis. One such example is ZEOMICS,[209] and this is an automated method that fully characterises the three-dimensional structure of zeolites. This uses the crystallographic coordinates of the material to identify all the portals, channels and cages, as well as the pore connectivity. It also calculates the pore size distribution, accessible volume, surface area, and largest cavity and pore limiting diameters.[209] Another high-throughput, geometry-based analysis tool, which can be used to understand the topology of crystalline porous materials, is Zeo^{++}.[210] This tool has been used extensively in this work, and is discussed in more detail in the subsequent chapters. Zeo^{++} has previously been used to map out and subsequently compare the pore topology of zeolites, with the view to separating ethane from ethene.[211,212]

1.9.5. Modelling on porous organic imine cages

Computational studies have also been carried out on the porous organic imine cages of Cooper and co-workers, to help rationalise their experimentally determined properties. The CSP for the crystal structure of two **CC1** polymorphs, **CC1α** and **CC1β**, have already been mentioned in Section 1.9.2. The interconversion of these polymorphs involves the conformer interconversion of **CC1**-*R* to **CC1**-*S*, and the barriers to rotations have been calculated using DFT calculations.[109] It was also possible to rationalise the selective porosity of these polymorphs using MD simulations and these calculations are the focus of Chapter 4.[104]

In addition, MC docking calculations (but not GCMC simulations) have been used to rationalise the nitrogen uptake of **CC2**, as the one-dimensional pore channel could not explain the total uptake. It was concluded that the intrinsic isolated cage voids seen in Figure 1.5d must also accessible to gas, and sorption locations for nitrogen for these sites, matched well with the electron density map determined from single X-ray diffractions patterns.[17] In the future, follow-up studies using MD simulations could be used to examine any diffusion pathways.

Molecular shape sorting has also been investigated in **CC3**.[213] The aim here was to study the selectivity of these cages to a number of guests (including toluene and xylene derivatives), using different computational techniques. First, gas

phase simulations were used to evaluate the shape and flexibility of the host, as well as the shape of the guest. The aim was to look at the mobility of guests in a single host molecule, as a computationally cheap indicator of the bulk mobility, and hence the selectivity. After this, solution-phase binding studies were used to examine the host-guest binding interactions and finally MD simulations were carried out in the crystalline lattice, to rationalise the diffusion of the guests. The combined results showed that the shape of the guest had a large role on its diffusivity through the diamondoid network; this is characterised by the restriction of mesitylene to a single cage during the MD simulations. This work was studied in conjunction with my work on smaller gases through **CC3**, detailed in Chapter 5.

Finally, for the large [8+12] cage, **CC7**, MD simulations showed that it was a lack of shape persistence in the individual molecules that led to the collapse of the structure and thus explaining experimental observations of an amorphous, non porous material upon desolvation.[107] This shows that by using gas phase MD simulations to represent the flexibility, systems that might initially be thought to be porous are in fact shown as non-porous, due to a lack of shape persistence.

1.10. Aim

Due to the unique chemical structure of cages **CC1**-**CC3**, no 'off-the-shelf' FF accurately described their dynamic motion, therefore it was decided to adapt a pre-existing FF, the PCFF, such that it contained the necessary parameters required to adequately simulate the cage structure. Using this bespoke force field, it was possible to use MD to examine the movement of gas through the cages crystal structure. This highlighted novel factors, such as pore 'hopping', window-to-channel diffusion, and window-to-window diffusion. Importantly, simulations could be compared to experimental observations, and this is especially important when considering the different polymorphic states of **CC1**. MD simulations were used to try and rationalise the observed 'on-off' porosity of the different polymorphs.

Also, as most organic molecules pack in such a way to minimise free space, they exhibit minimal void volume and hence permanent porosity in the solid state is rare. However, it has been shown experimentally that **CC3** is permanently porous to a variety of gases, although study of the static structure alone does not adequately explain the gas uptake. Therefore MD simulations were undertaken to try and rationalise this.

1.11. References

(1) Pierotti, R. A.; Rouquerol, J. Pure Appl Chem, **2006**, *57,* 1–17.

(2) Lloyd, G. O.; Barbour, L. J. Thesis: University of Stellenbosch. Dept. of Chemistry; Science, P. **2006**.

(3) Cheetham, A.; Férey, G.; Loiseau, T. Angew. Chem. Int. Ed. Engl., **1999**, *38*, 3268–3292.

(4) Côté, A. P. Science, **2005**, *310*, 1166–1170.

(5) Kitagawa, S.; Kitaura, R.; Noro, S. Angew. Chem. Int. Ed. **2004**, *43*, 2334–2375.

(6) Jones, J. T. A.; Hasell, T.; Wu, X.; Bacsa, J.; Jelfs, K. E.; Schmidtmann, M.; Chong, S. Y.; Adams, D. J.; Trewin, A.; Schiffman, F.; Cora, F.; Slater, B.; Steiner, A.; Day, G. M.; Cooper, A. I. Nature, **2011**, *474*, 367–371.

(7) Galarneau, A.; Cambon, H.; Di Renzo, F.; Fajula, F. Langmuir, **2001**, *17*, 8328–8335.

(8) Hernández, M. A.; Rojas, F.; Lara, V. H., J. Porous Mater. **2000**, *7*, 443–454.

(9) Horike, S.; Shimomura, S.; Kitagawa, S. Nature, **2009**, *1*, 695–704.

(10) Bradshaw, D.; Claridge, J. B.; Cussen, E. J.; Prior, T. J.; Rosseinsky, M. J. Acc. Chem. Res., **2005**, *38*, 273–282.

(11) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. Nature, **2003**, *423*, 705–714.

(12) Ananchenko, G. S.; Moudrakovski, I. L.; Coleman, A. W.; Ripmeester, J. A. Angew. Chem. Int. Ed. Engl. **2008**, *47*, 5616–5618.

(13) Cooper, A. I. Adv. Mater. **2009**, *21*, 1291–1295.

(14) Gutsche, C. D. Acc. Chem. Res. **1983**, *16*, 161–170.

(15) McKeown, N. B. J. Mater. Chem. **2010**, *20*, 10588.

(16) Lagona, J.; Mukhopadhyay, P. Angew. Chem. Int. Ed. **2005**, *44*, 4844-4870.

(17) Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. Nature Materials **2009**, *8*, 973–978.

(18) Mastalerz, M. Chem. Eur. J. **2012**, *18*, 10082-10091.

(19) Atwood, J. L. Science **2002**, *298*, 1000–1002.

(20) Pignataro, B. *Molecules at Work*, Wiley, **2012**.

(21) Barbour, L. J. Chem. Commun. **2006**, 1163-1168.

(22) Holst, J. R.; Trewin, A.; Cooper, A. I. Nat Chem **2010**, *2*, 915–920.

(23) Moulton, B.; Zaworotko, M. J. Chem. Rev. **2001**, *101*, 1629–1658.

(24) Hasell, T.; Chong, S. Y.; Jelfs, K. E.; Adams, D. J.; Cooper, A. I. J. Am. Chem. Soc. **2012**, *134*, 588–598.

(25) Jelfs, K. E.; Cooper, A. I. Curr. Opin. Solid State Mater. Sci. **2013**, *17*, 19–30.

(26) Jiang, S.; Jones, J. T. A.; Hasell, T.; Blythe, C. E.; Adams, D. J.; Trewin, A.; Cooper, A. I. Nat. Commun. **2011**, *2*, 207–7.

(27) Mastalerz, M. Chem. Eur. J. **2012**, *18*, 10082–10091.

(28) Colella, C.; Gualtieri, A. F. Microporous Mesoporous Mater. **2007**, *105*, 213–221.

(29) Rahman, M. M.; Hasnida, N.; Wan Nik, W. B. Jnl. Sci. Res. **2009**, *1*.

(30) http://www.bza.org/zeolites.html.

(31) Corma, A.; Garcia, H. Chem. Commun. **2004**, 1443-1459.

(32) Fraenkel, D.; Shabtai, J. J. Am. Chem. Soc. **1977**, *99*, 7074–7076.

(33) Lok, B. M.; Cannan, T. R.; Messina, C. A. Zeolites **1983**, *3*, 282–291.

(34) Davis, M. E.; Lobo, R. F. Chem. Mater. 1992, 4, 756–768.

(35) Dorfner, K. *Ion Exchangers*, Walter de Gruyter, **1991**.

(36) O'Keeffe, M.; Yaghi, O. M. Chem. Rev. **2012**, *112*, 675–702.

(37) Rowsell, J. L. C.; Yaghi, O. M. Microporous Mesoporous Mater. **2004**, *73*, 3–14.

(38) Eddaoudi, M.; Li, H.; Yaghi, O. M. J. Am. Chem. Soc. **2000**, *122*, 1391–1397.

(39) O'Keeffe, M. Chem. Soc. Rev. **2009**, *38*, 1215–1217.

(40) Eddaoudi, M. Science, **2002**, *295*, 469–472.

(41) Rosi, N. L. Science, **2003**, *300*, 1127–1129.

(42) Liu, J.; Thallapally, P. K.; McGrail, B. P.; Brown, D. R. Chem. Soc. Rev., **2012**, *41*, 2308–2322

(43) Gu, Z.-Y.; Jiang, D.-Q.; Wang, H.-F.; Cui, X.-Y.; Yan, X.-P. J. Phys. Chem. C. **2010**, *114*, 311–316.

(44) Alaerts, L.; Kirschhock, C. E. A.; Maes, M.; van der Veen, M. A.; Finsy, V.; Depla, A.; Martens, J. A.; Baron, G. V.; Jacobs, P. A.; Denayer, J. F. M.; De Vos, D. E. Angew. Chem. Int. Ed. Engl. **2007**, *46*, 4293–4297.

(45) Liu, J.; Thallapally, P. K.; Strachan, D. Langmuir **2012**, *28*, 11584–11589.

(46) Holst, J. R.; Cooper, A. I. Adv. Mater. **2010**, *22*, 5212–5216.

(47) Rose, M.; Böhlmann, W.; Sabo, M.; Kaskel, S. Chem. Commun. **2008**, 2462–2464.

(48) Feng, X.; Ding, X.; Jiang, D. Chem. Soc. Rev. **2012**, *41*, 6010–6022.

(49) Lukose, B.; Kuc, A.; Heine, T. Chem. Eur. J. **2011**, *17*, 2388-2392.

(50) Tilford, R. W.; Mugavero, S. J., III; Pellechia, P. J.; Lavigne, J. J. Adv. Mater. **2008**, *20*, 2741–2746.

(51) El-Kaderi, H. M.; Hunt, J. R.; Mendoza-Cortes, J. L.; Cote, A. P.; Taylor, R. E.; O'Keeffe, M.; Yaghi, O. M. Science, **2007**, *316*, 268–272.

(52) Han, S. S.; Furukawa, H.; Yaghi, O. M.; Goddard, W. A. J. Am. Chem. Soc. **2008**, *130*, 11580–11581.

(53) Furukawa, H.; Yaghi, O. M. J. Am. Chem. Soc. **2009**, *131*, 8875–8883.

(54) Jiang, J.-X.; Su, F.; Trewin, A.; Wood, C. D.; Campbell, N. L.; Niu, H.; Dickinson, C.; Ganin, A. Y.; Rosseinsky, M. J.; Khimyak, Y. Z.; Cooper, A. I. Angew. Chem. Int. Ed. Engl. **2007**, *46*, 8574–8578.

(55) Jiang, J.-X.; Su, F.; Trewin, A.; Wood, C. D.; Niu, H.; Jones, J. T. A.; Khimyak, Y. Z.; Cooper, A. I. J. Am. Chem. Soc. **2008**, *130*, 7710–7720.

(56) Jiang, J.-X.; Wang, C.; Laybourn, A.; Hasell, T.; Clowes, R.; Khimyak, Y. Z.; Xiao, J.; Higgins, S. J.; Adams, D. J.; Cooper, A. I. Angew. Chem. Int. Ed. **2010**, *50*, 1072–1075.

(57) Chen, L.; Honsho, Y.; Seki, S.; Jiang, D. J. Am. Chem. Soc. **2010**, *132*, 6742–6748.

(58) Chen, L.; Yang, Y.; Jiang, D. J. Am. Chem. Soc. **2010**, *132*, 9138–9143.

(59) McKeown, N. B.; Hanif, S.; Msayib, K.; Tattershall, C. E.; Budd, P. M. Chem. Commun. **2002**, 2782–2783.

(60) Budd, P. M.; Ghanem, B. S.; Makhseed, S. Chem. Commun. **2004**, 230–231.

(61) McKeown, N. B.; Makhseed, S.; Budd, P. M. Chem. Commun. **2002**, 2780–2781.

(62) McKeown, N. B.; Budd, P. M. Chem. Soc. Rev., **2006**, *35*, 675–683.

(63) Budd, P. M.; McKeown, N. B.; Fritsch, D. Macromol. Symp. **2006**, 245-246, 403-405.

(64) Ghanem, B. S.; Msayib, K. J.; McKeown, N. B.; Harris, K. D. M.; Pan, Z.; Budd, P. M.; Butler, A.; Selbie, J.; Book, D.; Walton, A. Chem. Commun. **2007**, 67–69.

(65) Budd, P. M.; Ghanem, B.; Msayib, K.; McKeown, N. B. J. Mater. Chem. **2003**, *13*, 2721–2726.

(66) Barrer, R. M.; Shanson, V. H. J. Chem. Soc., Chem. Commun. **1976**, 333.

(67) Flippen, J. L.; Karle, J.; Karle, I. L. J. Am. Chem. Soc. **1970**, *92*, 3749–3755.

(68) Ananchenko, G. S.; Udachin, K. A.; Pojarova, M. Cryst. Growth Des.

2006, *6*, 2141-2148.

(69) Atwood, J. L.; Barbour, L. J.; Jerga, A. Angew. Chem. Int. Ed. Engl. **2004**, *43*, 2948–2950.

(70) Gutsche, C. D.; Dhawan, B.; No, K. H.; Muthukrishnan, R. J. Am. Chem. Soc. **1981**, *103*, 3782–3792.

(71) Atwood, J. L. Science, **2002**, *296*, 2367–2369.

(72) Mastalerz, M.; Hüggenberg, W.; Dyker, G. Eur. J. Org. Chem. **2006**, 3977–3987.

(73) Isaacs, L. Chem. Commun. **2009**, 619–629.

(74) Kim, H.; Kim, Y.; Yoon, M.; Lim, S.; Park, S. M.; Seo, G.; Kim, K. J. Am. Chem. Soc. **2010**, *132*, 12200–12202.

(75) Lim, S.; Kim, H.; Selvapalam, N.; Kim, K.-J.; Cho, S. J.; Seo, G.; Kim, K. Angew. Chem. Int. Ed. Engl. **2008**, *47*, 3352–3355.

(76) Liu, L.; Zhao, N.; Scherman, O. A. Chem. Commun. **2008**, 1070–1072.

(77) Tian, J.; Ma, S.; Thallapally, P. K.; Fowler, D.; McGrail, B. P.; Atwood, J. L. Chem. Commun. **2011**, *47*, 7626.

(78) Tian, J.; Thallapally, P. K.; Dalgarno, S. J.; McGrail, P. B.; Atwood, J. L. Angew. Chem. **2009**, *121*, 5600–5603.

(79) Ulijn, R. V.; Smith, A. M. Chem. Soc. Rev. **2008**, *37*, 664.

(80) Comotti, A.; Bracco, S.; Distefano, G.; Sozzani, P. Chem. Commun. **2009**, 284–286.

(81) Soldatov, D. V.; Moudrakovski, I. L.; Grachev, E. V.; Ripmeester, J. A. J. Am. Chem. Soc. **2006**, *128*, 6737–6744.

(82) Bezzu, C. G.; Helliwell, M.; Warren, J. E.; Allan, D. R.; McKeown, N. B. Science, **2010**, *327*, 1627–1630.

(83) Msayib, K. J.; Book, D.; Budd, P. M.; Chaukura, N.; Harris, K. D. M.; Helliwell, M.; Tedds, S.; Walton, A.; Warren, J. E.; Xu, M.; McKeown, N. B. Angew. Chem. Int. Ed. **2009**, *48*, 3273–3277.

(84) Allcock, H. R.; Siegel, L. A. J. Am. Chem. Soc. **1964**, *86*, 5140–5144.

(85) Sozzani, P.; Comotti, A.; Bracco, S.; Simonutti, R. Angew. Chem. Int. Ed. Engl. **2004**, *43*, 2792–2797.

(86) Sozzani, P.; Bracco, S.; Comotti, A.; Ferretti, L.; Simonutti, R. Angew. Chem. Int. Ed. Engl. **2005**, *44*, 1816–1820.

(87) Hertzsch, T.; Budde, F.; Weber, E.; Hulliger, J. Angew. Chem. Int. Ed. Engl. **2002**, *41*, 2282–2284.

(88) Couderc, G.; Hertzsch, T.; Behrnd, N. R.; Krämer, K. Microporous Mesoporous Mater. **2006**, *88*, 170-175.

(89) Couderc, G.; Hulliger, J. Chem. Soc. Rev. **2010**, *39*, 1545–1554.

(90) Mastalerz, M.; Oppel, I. M. Angew. Chem. Int. Ed. **2012**, *51*, 5252–5255.
(91) Liu, X.; Liu, Y.; Li, G.; Warmuth, R. Angew. Chem. Int. Ed. **2006**, *45*, 901–904.
(92) Liu, X.; Warmuth, R. J. Am. Chem. Soc. **2006**, *128*, 14120–14127.
(93) Liu, X.; Warmuth, R. Nat Protoc. **2007**, *2*, 1288–1296.
(94) Liu, X.; Liu, Y.; Warmuth, R. Supramol. Chem. **2008**, *20*, 41–50.
(95) Mastalerz, M. Chem. Commun. **2008**, 4756.
(96) Mastalerz, M.; Schneider, M. W.; Oppel, I. M.; Presly, O. Angew. Chem. Int. Ed. **2010**, *50*, 1046–1051.
(97) Schneider, M. W.; Oppel, I. M.; Ott, H.; Lechner, L. G.; Hauswald, H.-J. S.; Stoll, R.; Mastalerz, M. Chem. Eur. J. **2011**, *18*, 836–847.
(98) Schneider, M. W.; Lechner, L. G.; Mastalerz, M. J. Mater. Chem. **2012**, *22*, 7113–7116.
(99) Schneider, M. W.; Oppel, I. M.; Mastalerz, M. Chem. Eur. J. **2012**, *18*, 4156–4160.
(100) Jin, Y.; Voss, B. A.; Noble, R. D.; Zhang, W. Angew. Chem. **2010**, *122*, 6492–6495.
(101) Jin, Y.; Voss, B. A.; Jin, A.; Long, H.; Noble, R. D.; Zhang, W. J. Am. Chem. Soc. **2011**, *133*, 6650–6658.
(102) Jiang, S.; Bacsa, J.; Wu, X.; Jones, J. T. A.; Dawson, R.; Trewin, A.; Adams, D. J.; Cooper, A. I. Chem. Commun. **2011**, *47*, 8919.
(103) Jones, J. T. A.; Holden, D.; Mitra, T.; Hasell, T.; Adams, D. J.; Jelfs, K. E.; Trewin, A.; Willock, D. J.; Day, G. M.; Bacsa, J.; Steiner, A.; Cooper, A. I. Angew. Chem. Int. Ed. Engl. **2011**, *50*, 749–753.
(104) Jones, J. T. A.; Holden, D.; Mitra, T.; Hasell, T.; Adams, D. J.; Jelfs, K. E.; Trewin, A.; Willock, D. J.; Day, G. M.; Bacsa, J.; Steiner, A.; Cooper, A. I. Angew. Chem. Int. Ed. Engl. **2011**, *50*, 749–753.
(105) Jiang, S.; Bacsa, J.; Wu, X.; Jones, J. T. A.; Dawson, R.; Trewin, A.; Adams, D. J.; Cooper, A. I. Chem. Commun. **2011**, *47*, 8919.
(106) Mitra, T.; Wu, X.; Clowes, R.; Jones, J. T. A.; Jelfs, K. E.; Adams, D. J.; Trewin, A.; Bacsa, J.; Steiner, A.; Cooper, A. I. Chem. Eur. J. **2011**, *17*, 10235–10240.
(107) Jelfs, K. E.; Wu, X.; Schmidtmann, M.; Jones, J. T. A.; Warren, J. E.; Adams, D. J.; Cooper, A. I. Angew. Chem. Int. Ed. **2011**, *50*, 10653–10656.
(108) Hasell, T.; Wu, X.; Jones, J. T. A.; Bacsa, J.; Steiner, A.; Mitra, T.; Trewin, A.; Adams, D. J.; Cooper, A. I. Nat. Chem. **2010**, *2*, 750–755.
(109) Jelfs, K. E.; Schiffmann, F.; Jones, J. T. A.; Slater, B.; Cora, F.; Cooper,

A. I. Phys. Chem. Chem. Phys. **2011**, *13*, 20081–20085.

(110) Yaghi, O. M.; Li, H.; Eddaoudi, M.; O'Keeffe, M. Nature, **1999**, *402*, 276–279.

(111) Ben, T.; Ren, H.; Ma, S.; Cao, D.; Lan, J.; Jing, X.; Wang, W.; Xu, J.; Deng, F.; Simmons, J. M.; Qiu, S.; Zhu, G. Angew. Chem. Int. Ed. **2009**, *48*, 9457–9460.

(112) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, A. O.; Snurr, R. Q.; O'Keeffe, M.; Kim, J.; Yaghi, O. M. Science, **2010**, *329*, 424–428.

(113) Mastalerz, M.; Schneider, M. W. Chem. Commun. **2012**, *48*, 9861-9863.

(114) Hasell, T.; Chong, S. Y.; Schmidtmann, M.; Adams, D. J.; Cooper, A. I. Angew. Chem. Int. Ed. Engl. **2012**, *51*, 7154–7157.

(115) Hansen, J.; Sato, M.; Kharecha, P.; Beerling, D.; Berner, R.; Masson-Delmotte, V.; Pagani, M.; Raymo, M.; Royer, D. L.; Zachos, J. C. Open Atmosp. Sci. J. **2008**, *2*, 217-231.

(116) Babarao, R.; Eddaoudi, M.; Jiang, J. W. Langmuir **2010**, *26(13)*, 11196-11203.

(117) Heinberg, R. *The Party's Over: Oil, War and the fate of Industrial Societies*, New Society Publishers, **2005**.

(118) Hufton, J. R.; Mayorga, S.; Sircar, S. AIChE J. **1999**, *2*, 248-256.

(119) Vitillo, J. G. J.; Ricchiardi, G. G.; Spoto, G. G.; Zecchina, A. A. Phys. Chem. Chem. Phys. **2005**, *7*, 3948–3954.

(120) Weitkamp, J. Int. J. Hydrogen Energy, **1995**, *20*, 967–970.

(121) European Patent App. EP1188477, **2002**.

(122) http://www.hydrogen.energy.gov/storage.html.

(123) Furukawa, H.; Miller, M. A.; Yaghi, O. M. J. Mater. Chem. **2007**, *17*, 3197.

(124) Yan, Y.; Lin, X.; Yang, S.; Blake, A. J.; Dailly, A.; Champness, N. R.; Hubberstey, P.; Schröder, M. Chem. Commun. **2009**, 1025–1027.

(125) Getman, R. B.; Bae, Y. S.; Wilmer, C. E.; Snurr, R. Q. Chem. Rev. **2011**, *112*, 703-723.

(126) Jensen, N. K.; Rufford, T. E.; Watson, G.; Zhang, D. K.; Chan, K. I.; May, E. F. J. Chem. Eng. Data. **2012**, *57*, 106–113.

(127) Kidnay, A. J.; Parrish, W. R. *Fundamentals of Natural Gas Processing,* CRC Press, **2006**.

(128) Thome, J. R. Wolverine Tube Inc, **2004**.

(129) Ma, L.; Lee, J. Y.; Li, J.; Lin, W. Inorg. Chem. **2008**, *47*, 3955–3957.

(130) Sun, D.; Ma, S.; Simmons, J. M.; Li, J.-R.; Yuan, D.; Zhou, H.-C. Chem.

Commun. **2010**, *46*, 1329–1331.

(131) Morris, R. E.; Wheatley, P. S. Angew. Chem. Int. Ed. Engl. **2008**, *47*, 4966–4981.

(132) Millward, A. R.; Yaghi, O. M. J. Am. Chem. Soc. **2005**, *127*, 17998–17999.

(133) Dorcheh, A. S.; Denysenko, D.; Volkmer, D.; Donner, W.; Hirscher, M. Microporous and Mesoporous Materials **2012**, *162*, 64–68.

(134) Cullen, S. C.; Gross, E. G. Science, **1951**, *113*, 580–582.

(135) Weiskopf, R. B.; Lynch, C., III; Baum, J.; Tenbrinck, R. Anesthesiology 2000, *92*, 865-868.

(136) Duhamel, G.; Choquet, P.; Grillon, E.; Lamalle, L.; Leviel, J.-L.; Ziegler, A.; Constantinesco, A. Magn. Reson. Med. **2001**, *46*, 208–212.

(137) Ryan, P.; Farha, O. K.; Broadbelt, L. J.; Snurr, R. Q. AIChE J. **2010**, *57*, 1759–1766.

(138) Kerry, F. G. *Industrial Gas Handbook: Gas Separation and Purification*, CRC Press, **2010**.

(139) Thallapally, P. K.; Grate, J. W.; Motkuri, R. K. Chem. Commun. **2012**, *48*, 347–349.

(140) Fernandez, C. A.; Liu, J.; Thallapally, P. K.; Strachan, D. M. J. Am. Chem. Soc. **2012**, *134*, 9046−9049

(141) Munakata, K.; Fukumatsu, T. J. Nucl. Sci. Technol. **1999**, *36*, 818-829.

(142) Jameson, C. J.; Jameson, A. K.; Lim, H.-M. J. Chem. Phys. **1997**, *107*, 4364.

(143) Potapov, S. V.; Fomkin, A. A.; Sinitsyn, V. A.; Shkolin, A. V. Prot Met Phys Chem Surf. **2010**, *46*, 639–643.

(144) Voth, G. A. *Coarse-Graining of Condensed Phase and Bimolecular Systems*, CRC Press, **2008**.

(145) Harmandaris, V. A.; Adhikari, N. P.; Van der Vegt, N. Macromolecules. **2006**, *39*, 6708-6719.

(146) Nemkevich, A.; Bürgi, H.-B.; Spackman, M. A.; Corry, B. Phys. Chem. Chem. Phys. **2010**, *12*, 14916.

(147) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. J. Am. Chem. Soc. **2001**, *114*, 10024–10035.

(148) Sun, H. J. Phys. Chem. B **1998**, *102*, 7338–7364.

(149) Sun, H. Macromolecules. **1995**, *28*, 701-712.

(150) Yang, J.; Ren, Y.; Tian, A.-M.; Sun, H. J. Phys. Chem. B **2000**, *104*, 4951–4957.

(151) Sun, H.; Mumby, S. J.; Maple, J. R. *J. Am. Chem. Soc.* ***1994,*** 116***,*** 2918-

2981.

(152) Sun, H. J. Comput. Chem. **1994**, *15*, 752–768.

(153) Hill, J. R.; Sauer, J. J. Phys. Chem. **1994**, *98*, 1238–1244.

(154) Schmid, R.; Tafipolsky, M. J. Am. Chem. Soc. **2008**, *130*, 12600–12601.

(155) Tafipolsky, M.; Schmid, R. J. Phys. Chem. B **2009**, *113*, 1341–1352.

(156) Huang, B. L.; Ni, Z.; Millward, A.; McGaughey, A. J. H.; Uher, C.; Kaviany, M.; Yaghi, O. Int. J. Heat Mass Transfer. **2007**, *50*, 405–411.

(157) Greathouse, J. A.; Allendorf, M. D. J. Phys. Chem. C **2008**, *112*, 5795–5802.

(158) Zheng, B.; Sant, M.; Demontis, P.; Suffritti, G. B. J. Phys. Chem. C. **2012**, *116*, 933–938.

(159) Holden, D.; Jelfs, K. E.; Cooper, A. I.; Trewin, A.; Willock, D. J. J. Phys. Chem. C. **2012**, *116*, 16639–16651.

(160) Price, S. L. Acta Cryst. **2013**. *B69*, 313–328.

(161) Zwijnenburg, M. A.; Illas, F.; Bromley, S. T. Phys. Rev. Lett. **2010**, *104*, 175503.

(162) Zwijnenburg, M. A.; Jelfs, K. E.; Bromley, S. T. Phys. Chem. Chem. Phys. **2010**, *12*, 8505–8512.

(163) Foster, M. D.; Simperler, A.; Bell, R. G.; Friedrichs, O. D.; Paz, F. A. A.; Klinowski, J. Nat. Mat. **2004**, *3*, 234–238.

(164) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Nat. Chem. **2012**, *4*, 83-89.

(165) Mellot-Draznieks, C.; Newsam, J.; Gorman, A.; Freeman, C.; Férey, G. Angew. Chem. Int. Ed. Engl. **2000**, *39*, 2270–2275.

(166) Zwijnenburg, M. A.; Bromley, S. T.; Foster, M. D.; Bell, R. G.; Delgado-Friedrichs, O.; Jansen, J. C.; Maschmeyer, T. Chem. Mater. **2004**, *16*, 3809–3820.

(167) Price, S. Acc. Chem. Res. **2008**, *42*, 117-126.

(168) Neumann, M. A.; Perrin, M.-A. J. Phys. Chem. B **2005**, *109*, 15531–15541.

(169) Cruz-Cabeza, A. J.; Day, G. M.; Jones, W. Chemistry **2009**, *15*, 13033–13040.

(170) Sarkisov, L. Adv. Mater. **2012**, *24*, 3130–3133.

(171) Abbott, L. J.; McKeown, N. B.; Colina, C. M. J. J. Mater. Chem. A, **2013**, *1*, 11950–11960.

(172) Abbott, L. J.; Colina, C. M. Macromolecules. **2011**, *44*, 4511–4519.

(173) Abbott, L. J.; McDermott, A. G.; Del Regno, A.; Taylor, R. G. D.; Bezzu, C. G.; Msayib, K. J.; McKeown, N. B.; Siperstein, F. R.; Runt, J.; Colina,

C. M. J. Phys. Chem. B. **2013**, *117*, 355–364.

(174) Smit, B.; Krishna, R. Curr. Opin. Solid State Mater. Sci. **2001**, *5*, 455–461.

(175) Krishna, R.; van Baten, J. M. Phys. Chem. Chem. Phys. **2011**, *13*, 10593–10616.

(176) Keskin, S.; Liu, J.; Rankin, R. B.; Johnson, J. K.; Sholl, D. S. Ind. Eng. Chem. Res. **2009**, *48*, 2355–2371.

(177) Salles, F.; Bourrelly, S.; Jobic, H.; Devic, T.; Guillerm, V.; Llewellyn, P.; Serre, C.; Férey, G.; Maurin, G. J. Phys. Chem. C. **2011**, *115*, 10764–10776.

(178) Daschbach, J. L.; Sun, X.; Thallapally, P. K.; McGrail, B. P.; Dang, L. X. J. Phys. Chem. B. **2010**, *114*, 5764–5768.

(179) Williams, J. J.; Wiersum, A. D.; Seaton, N. A.; Düren, T. J. Phys. Chem. C. **2010**, *114*, 18538–18547.

(180) Coudert, F.-X.; Mellot-Draznieks, C.; Fuchs, A. H.; Boutin, A. J. Am. Chem. Soc. **2009**, *131*, 11329–11331.

(181) Neimark, A. V.; Coudert, F.-X.; Boutin, A.; Fuchs, A. H. J. Phys. Chem. Lett. **2010**, *1*, 445–449.

(182) Serre, C.; Mellot-Draznieks, C.; Surblé, S.; Audebrand, N.; Filinchuk, Y.; Férey, G. Science. **2007**, *315*, 1828–1831.

(183) Moggach, S. A.; Bennett, T. D.; Cheetham, A. K. Angew. Chem. Int. Ed. Engl. **2009**, *48*, 7087–7089.

(184) Düren, T.; Sarkisov, L.; Yaghi, O. M.; Snurr, R. Q. Langmuir. **2004**, *20*, 2683–2689.

(185) Wang, S. Energy & Fuels. **2007,** *21,* 953-956.

(186) Gallo, M.; Glossman-Mitnik, D. J. Phys. Chem. C. **2009**, *113*, 6634-6642.

(187) Sirjoosingh, A.; Alavi, S.; Woo, T. K. J. Phys. Chem. **2010**, *114*, 2171-2178.

(188) Coasne, B.; Galarneau, A.; Di Renzo, F.; Pellenq, R. J. M. Langmuir. **2010**, *26*, 10872–10881.

(189) Demontis, P.; Suffritti, G. B. Chem. Rev. **1997**, *97*, 2845–2878.

(190) Jiang, J.-X.; Su, F.; Niu, H.; Wood, C. D.; Campbell, N. L.; Khimyak, Y. Z.; Cooper, A. I. Chem. Commun. **2008**, 486–488.

(191) Dubbeldam, D.; Krishna, R.; Snurr, R. Q. J. Phys. Chem. **2009**, *133*, 19317-19327.

(192) Haldoupis, E.; Watanabe, T.; Nair, S.; Sholl, D. S. ChemPhysChem. **2012**, *13*, 3449–3452.

(193) Zhao, L.; Yang, Q.; Ma, Q.; Zhong, C.; Mi, J.; Liu, D. J Mol Model. **2011**,

17, 227–234.

(194) Avellaneda, A.; Valente, P.; Burgun, A.; Evans, J. D.; Markwell-Heys, A. W.; Rankine, D.; Nielsen, D. J.; Hill, M. R.; Sumby, C. J.; Doonan, C. J. Angew. Chem. Int. Ed. **2013**, *52*, 1-5.

(195) Xue, C.; Zhong, C. Chin. J. Chem. **2009**, *27*, 472-478.

(196) Smit, B.; Maesen, T. L. M. Nature. **2008**, *451*, 671–678.

(197) Combariza, A. F.; Sastre, G.; Corma, A. J. Phys. Chem. C. **2009**, *113*, 11246–11253.

(198) Brockway, A. M.; Schrier, J. J. Phys. Chem. C. **2013**, *117*, 393–402.

(199) Nemkevich, A.; Corry, B.; Spackman, M. A. Phys. Chem. Chem. Phys. **2012**, *14*, 1570.

(200) Adams, J. E.; Cox, J. R.; Christiano, A. J.; Deakyne, C. A. J. Phys. Chem. A. **2008**, *112*, 6829–6839.

(201) Krishna, R. J. Phys. Chem. C. **2009**, *113*, 19756–19781.

(202) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L.; Snurr, R. Q. Chem. Sci. **2012**, *3*, 2217.

(203) Bae, Y.-S.; Hauser, B. G.; Colón, Y. J.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q. Microporous Mesoporous Mater. **2013**, *169*, 176–179.

(204) Haldoupis, E.; Nair, S.; Sholl, D. S. J. Am. Chem. Soc. **2010**, *132*, 7528–7539.

(205) Haldoupis, E.; Nair, S.; Sholl, D. S. Phys. Chem. Chem. Phys. **2011**, *13*, 5053-5060.

(206) Haldoupis, E.; Nair, S.; Sholl, D. S. J. Am. Chem. Soc. **2012**, *134*, 4313–4323.

(207) Van Heest, T.; Teich-McGoldrick, S. L.; Greathouse, J. A.; Allendorf, M. D.; Sholl, D. S. J. Phys. Chem. C. **2012**, *116*, 13183–13195.

(208) First, E. L.; Gounaris, C. E.; Wei, J.; Floudas, C. A. Phys. Chem. Chem. Phys. **2011**, *13*, 17339–17358.

(209) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Microporous Mesoporous Mater. **2012**, *149*, 134–141.

(210) Kim, J.; Lin, L.-C.; Martin, R. L.; Swisher, J. A.; Haranczyk, M.; Smit, B. Langmuir. **2012**, *28*, 11914–11919.

(211) Pinheiro, M.; Martin, R. L.; Rycroft, C. H.; Jones, A.; Iglesia, E.; Haranczyk, M. J. Mol. Graphics Modell. **2013**, *44*, 208–219.

(212) Mitra, T.; Jelfs, K. E.; Schmidtmann, M.; Ahmed, A.; Chong, S. Y.; Adams, D. J.; Cooper, A. I. Nat. Chem. 2013, ***5***, 276-281.

Chapter 2

Experimental and theoretical methodology

2.1. Experimental methodology

2.1.1. Single crystal X-ray diffraction

This thesis concerns crystalline porous solids. One of the best ways to identify the structure of an ordered material is to use single crystal X-Ray diffraction. The atoms within a crystal align with a characteristic interatomic distance; when the wavelength of the X-rays is similar to this distance, diffraction can occur. The distance between the atomic planes, *d*, can be calculated by using the information of the path lengths (*AB* and *BC*) and θ, which is the angle at that the wavelength both enters and leaves the atomic planes. This allows *d* to be calculated by:

$$\text{Path Length Difference} = AB + BC = 2\,d\sin\theta \qquad \text{(Eq. 2.1)}$$

Using Bragg's law, the wavelength of the X-ray can be computed:

$$n\lambda = 2\,d\sin\theta \qquad \text{(Eq. 2.2)}$$

Here, *n* is an integer and λ is the wavelengths of the X-rays. As the X-rays enter the sample, they are diffracted back by the atomic planes of the crystal. This generates diffraction patterns at specific angles that can in turn be interpreted using Bragg's Law (Equation 2.2). This process is shown schematically in Figure 2.1.

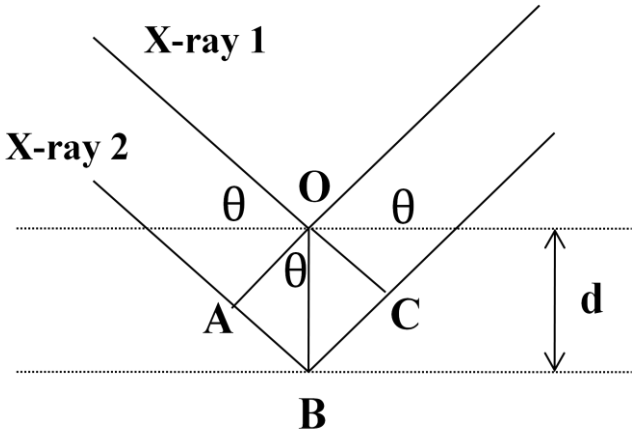


Figure 2.1. Figure showing Bragg's law of diffraction where two X-rays enter a crystal plane with a spacing, d. This was taken from reference. [1]

This technique provides a wide range of information for the system, including the unit cell parameters, position of atoms, and chemical bonds within the lattice.[2] To obtain this information the sample has to be of sufficient size and quality. This technique works by the generation of a diffraction pattern from a single crystal –

this gives an array of spots in reciprocal space. From this the direction *hkl* reflections within the unit cell can be determined; these relate to the structure factor, F_{hkl}. A Fourier Transformation is then applied to generate the crystal structure of the system. All data for this work was collected and refined by Dr John Basca, University of Liverpool. From a modelling perspective, one point to note is that hydrogen atoms are not located directly but rather by using standard bond distances. Also, X-ray diffraction is a time-averaging technique that gives an essentially 'static' snapshot of the structure of a given material.

2.1.2. Gas adsorption

The introduction of gas molecules to a surface is known as gas adsorption; this is illustrated in Figure 2.2. Adsorption is the binding of a molecule to a surface and can be either chemical (chemisorption)[3] or physical (physisorption). Chemisorption requires the formation of a chemical bond between the surface and the absorbent, whereas physisorption involves a much weaker interaction, and instead of an electron transfer between the surface and the absorbent, it is due to a polarization of the two substituents or weak van der Waals interactions.[4] These two types of adsorptions are compared in Table 2.1. A good example of materials that exhibit chemisorption are metal hydrides.[5,6] The advantages of chemisorptive materials are their relatively high binding energies; however, high operation temperatures for desorption tend to hinder their use in gas storage, for example.

Chemisorption	**Physisorption**
Electron exchange occurs	Polarization occurs
Formation of a chemical bond	Formation of van der Waals interactions
Strong	Weak
Binding Energy ≥ 1 eV (100 kJ mol^{-1})	Binding Energy ≤ 0.3 eV (30 kJ mol^{-1}) This makes them only stable at cryogenic temperatures
Analogies with coordination chemistry due to strength of binding energy	Less strongly directional

Table 2.1. Comparison between chemisorption and physisorption. Adapted from reference 3.

The uptake of hydrogen by physisorption in high surface area materials could be beneficial for energy storage purposes. Physisorption is dictated by van der Waals interactions between the adsorbate and a surface. As no chemical bonds are formed, the adsorption heat for physisorption is lower than chemisorption. Since no permanent bonds are formed, physisorption is more easily reversible. This is particularly advantageous when attempting to design gas storage materials. If

the pores are of sufficient size, multiple layers of adsorbates can also form on top of one another, generating multilayers. With the application of appropriate theories, this can be used to estimate the surface area, pore size, and pore volume of the system in question.

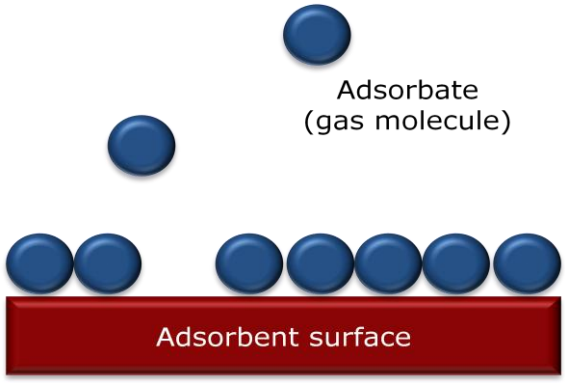


Figure 2.2. Scheme showing the adsorption of a gas onto the surface of an adsorbent.[7]

2.1.3. Gas sorption isotherms

A gas sorption isotherm is a plot of the amount of gas adsorbed, as a function of pressure, at a fixed temperature. To measure a desorption isotherm, uptakes are measured, as the pressure is reduced in a stepwise manner. Sorption isotherms can be split into six different classifications; these are illustrated in Figure 2.3. These classes are described as follows:

- Type I isotherms are seen predominantly for systems that are microporous in nature (pores < 2 nm); the majority of the adsorption will occur at relatively low pressures. Adsorption is often saturated after a relative pressure of around 0.5 or so.
- Type II isotherms are generally observed for non-porous solids. For these a monolayer on the surface is first obtained, followed by multilayer at higher partial pressures. This change is seen by an inflection on the graph.
- Type III isotherms are convex in nature relative to the pressure axis. This is generally found in systems where the adsorbate-adsorbent interactions are very weak; these can be both non-porous or microporous systems. The weak interactions lead to low uptake at low partial pressure. However, when a monolayer is established, the adsorbent is now acting as the surface and the adsorbent-adsorbent interactions are much stronger, thus leading to an accelerated uptake at higher pressure.
- Type IV isotherms generally occur for mesoporous systems. The shape of these isotherms is system specific. Capillary condensation can lead to a

hysteresis loop whereby the desorption isotherm deviates from the adsorption isotherm.

- Type V isotherms can be seen for both microporous and mesoporous systems. Here, like Type III isotherms, the uptakes are convex to the relative pressure axis. There is a hysteresis loop as the relative pressure is increased.
- Type VI isotherms are generated by the step-by-step formation of a complete monolayer followed by the adsorption of another complete layer and this is repeated as the relative pressure is increased. This is more of a hypothetical isotherm and is meant to represent what would be seen if the system were both extremely homogeneous as well as non-porous – real experimental type VI isotherms are rare.

There are several ways to model adsorption isotherms; these include the Langmuir adsorption model and the Brunauer-Emmett-Teller model, which are discussed below.

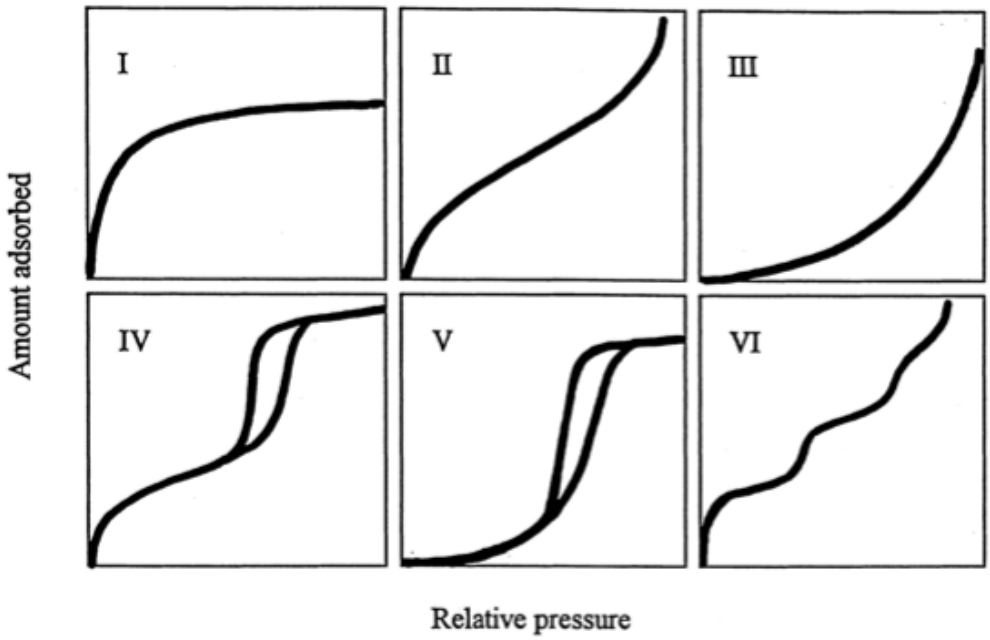


Figure 2.3. This figure shows the IUPAC classification of adsorption isotherms, taken from reference 8.

2.1.4. Langmuir adsorption model

The Langmuir adsorption model was first suggested in 1916 by Irving Langmuir[9] and is based on a number of assumptions:

1. Adsorbates are adsorbed as a monolayer on the sorbent only.
2. The sorbent surface is uniform and all sites on the surface are equivalent; these sites can only accommodate one adsorbate.

3. When the temperature is kept constant, there is a dynamic equilibrium between the gas and the adsorbed layer.
4. Once adsorbed, the adsorbates are localised and therefore do not migrate to neighbouring sites meaning that the enthalpy of adsorption for each site remains at a constant, irrespective of surface coverage.
5. Interactions between neighbouring adsorbates are ignored.
6. If the adsorbates collide with a vacant site on the adsorbent a bond is formed and the adsorbate remains on the surface. If however the site is filled, the adsorbate is deflected back up into the gas phase.

Langmuir theory dictates that when the pressure and temperature of the system are kept at a constant, there is a dynamic equilibrium between the rate of molecules sticking to the surface, the adsorption, and the rate of molecules leaving the surface, the desorption. This can be expressed as:

$$M + S \rightleftharpoons M\text{-}S \qquad \text{(Eq. 2.3)}$$

Where *M* is the molecules and *S* is the adsorbent site surface. This means that the rate for adsorption/desorption, *K*, can be expressed as:

$$K = k_{ads} / k_{des} \qquad \text{(Eq. 2.4)}$$

Where k_{ads} and k_{des} are the rate constants for the adsorption and desorption steps respectively. The total coverage of the adsorbent surface, θ, is expressed as a fraction deduced by the total number of occupied sites divided by the overall number of sites, therefore can never exceed 1:

$$\theta = N_{occupied} / N_{total} \qquad \text{(Eq. 2.5)}$$

The rate of adsorption is dependent on the pressure of the adsorbate, *P*, in combination with θ. This rate of adsorption can be calculated using:

$$\text{Rate of adsorption} = k_{ads}P(1 - \theta) \qquad \text{(Eq. 2.6)}$$

Where k_{ads} is the rate constant for adsorption step. To calculate the rate of desorption, the following expression is used:

$$\text{Rate of desorption} = k_{des}\theta \qquad \text{(Eq. 2.7)}$$

Although *P* is not explicitly required to calculate the rate of desorption, it does depend on θ; as *P* is needed to deduce θ there is an indirect influence. When the system is at equilibrium, the rate of adsorption and desorption are equal, meaning that:

$$k_{ads}P(1 - \theta) = k_{des}\theta \quad \text{(Eq. 2.8)}$$

As equation 2.4. shows the relationship between K_{ads} and K_{des}, Equation 2.8. can be subsequently arranged, so that:

$$K = \theta/P(1 - \theta) \quad \text{(Eq. 2.9)}$$

Alternatively, θ can be expressed as:

$$\theta = KP/1+KP \quad \text{(Eq. 2.10)}$$

Equation 2.10. is known as the Langmuir equation for associative adsorption. This can be used to predict what affect altering the pressure has on θ.

Another way to express θ is in terms of relative masses and volumes. This is useful as an estimate for the surface area can be then calculated; θ can be expressed as the total volume of gas adsorbed, V_{ads}, divided by the volume of the monolayer coverage, V_{mono}:

$$\theta = V_{ads}/V_{mono} \quad \text{(Eq. 2.11)}$$

From this, the Langmuir equation can now be rearranged in the form of a linear equation, such that:

$$P/V_{ads} = (1/KV_{mono}) + (P/V_{mono}) \quad \text{(Eq. 2.12)}$$

This allows a straight-line graph to be plotted, with P / V_{ads} against P; from this the surface area, SA_{lang}, can be determined:

$$SA_{lang} = V_{mono}\, La_m \quad \text{(Eq. 2.13)}$$

Where a_m is the molecular cross sectional area of the adsorbate (nm^2) and L is Avogadro's constant (6.023×10^{23} mol^{-1}).

2.1.5. BET model

The technique used most widely to describe adsorption isotherms is the Brunauer-Emmett-Teller (BET) model.[2] This is an extension of the Langmuir model which tries to include the formation of multilayers of adsorbates. One key similarity with the Langmuir model is that the BET model also assumes that the adsorbent surface is uniform in nature. On top of this there are a number of other assumptions made by the BET model:

1. Although the first monolayer has a fixed heat of adsorption, any subsequent layers have heats of adsorption equal to the latent heat of

evaporation.

2. At equilibrium, the rate of condensation and evaporation is balanced.

As the Langmuir model is applied to each monolayer, the BET equation has the following expression:

$$\frac{P}{V(P_0 - P)} = \frac{1}{v_m c} + \frac{(c-1)}{v_m c}\frac{P}{P_0} \qquad \text{(Eq. 2.14)}$$

Here P is the equilibrium pressure and P_0 is the saturated gas pressure at 77 K. V is the total gas adsorbed volume, whilst V_m is the volume of gas adsorbed on each monolayer; c is a constant.

From this, a linear graph can be obtained by plotting:

$$\frac{P}{V(P_0 - P)} \quad \text{against...} \quad \frac{P}{P_0} \qquad \text{(Eq. 2.15)}$$

This gives a gradient and intercept of:

$$\frac{(c-1)}{v_m c} \quad \text{and...} \quad \frac{1}{v_m c} \qquad \text{(Eq. 2.16)}$$

The BET model is frequently used for nitrogen adsorption isotherms at 77 K. This is because as both the gradient and intercept allow V_m to be calculated, it is possible to estimate the overall surface area, SA_{BET}, using the following equation:

$$SA_{BET} = \sigma V_m L / V_0 \qquad \text{(Eq. 2.17)}$$

Where V_0 is the molar volume of nitrogen, σ is the area occupied by each adsorbate molecule and L is Avogadro's constant. V_0 for nitrogen at standard temperature and pressure is 22,414 $dm^3\ mol^{-1}$ while σ for nitrogen is 16.2 $Å^2$.

2.1.6. Gas diffusion

Diffusion of gases through porous solids can play a pivotal role in a whole range of applications; indeed, diffusion can often be the rate-limiting step in such processes, especially catalytic activity.[10] The rate of diffusion is also critical in determining the selective properties of a system, for example, in molecular separations. The pore aperture within microporous materials is often comparable with the size of common gas molecules and therefore the relative size and shape of the pore topology has a large effect on the diffusion rates of the gases. There are several different ways to describe the diffusion rate of a gas at a

concentration, c – the three most common are the self-diffusivity, $D_s(c)$, the transport diffusivity $D_t(c)$, and the corrected diffusivity, $D_0(c)$. The difference between $D_s(c)$ and $D_t(c)$ is shown in Figure 2.4. $D_s(c)$ determines the displacement of a specific gas molecule as it diffuses; the system here is assumed to be at equilibrium. This can be defined using the Einstein equation:[11]

$$D_s(c) = \lim_{t\to\infty} \frac{1}{6t} \left\langle \left| \vec{r}(t) - \vec{r}(0) \right|^2 \right\rangle \qquad \text{(Eq. 2.18)}$$

Where $\vec{r}(t)$ is the position of the specific gas molecule at time, t. This helps calculate the diffusion of a three dimensional isotropic system, but can be generalised to anisotropic systems.

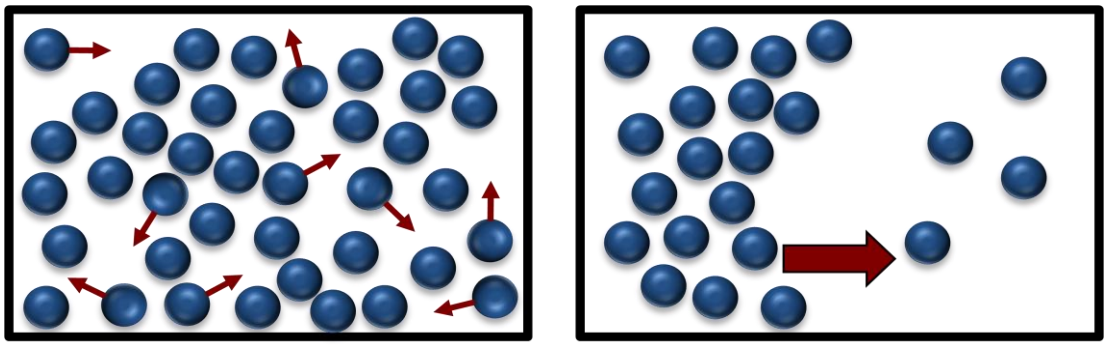


Figure 2.4. Scheme showing the difference between self-diffusivity (left) and transport diffusivity (right).[7]

As the self-diffusivity can, in theory, only be calculated accurately when t tends to ∞, the longer the simulation, the more accurate the results; this needs to be balanced of course with computational expense. To characterise the diffusion of a molecule in a macroscopic system, the property $D_t(c)$ needs to be calculated:[12]

$$\vec{J} = -D_t(c)\nabla c \qquad \text{(Eq. 2.19)}$$

This relates the macroscopic flux, $\vec{J}$, to the macroscopic concentration gradient, ∇c. Like $D_s(c)$, although $D_t(c)$ is suitable for isotropic systems, it can be generalised to anisotropic systems. Figure 2.4. illustrates $D_t(c)$ in more detail. When calculating $D_t(c)$, it is more convenient to represent this quantity by referring to the chemical potential of a system, as this is a driving force for diffusion in microporous materials:

$$D_t(c) = D_0(c)\left[\frac{\partial \ln f}{\partial \ln c}\right]_T \qquad \text{(Eq. 2.20)}$$

Where *f* is the fugacity of the bulk phase at equilibrium.[10] All three of these diffusivities are concentration dependent, and therefore are only equivalent when the concentration is zero:[12]

$$D_s(c=0) = D_t(c=0) = D(0)(c=0) = D(0) \qquad \text{(Eq. 2.21)}$$

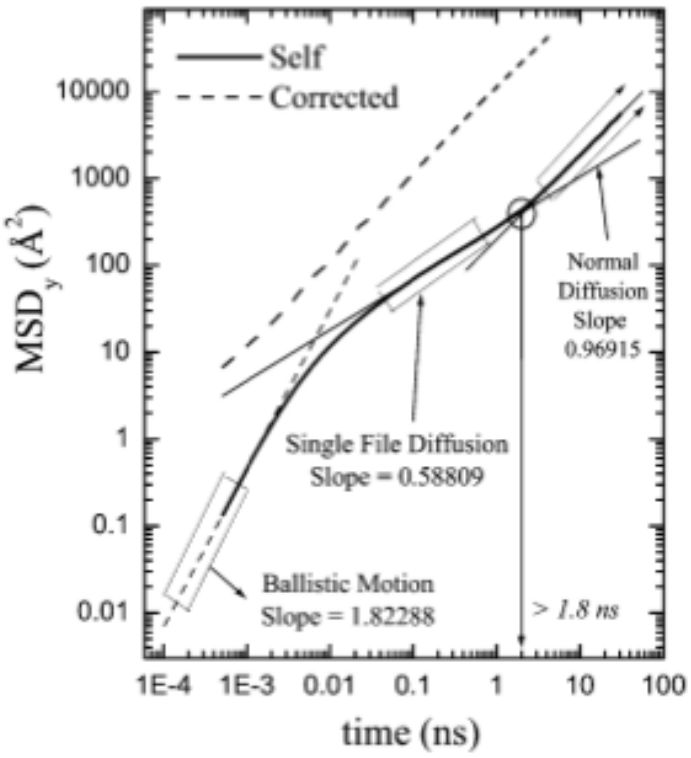


Figure 2.5. Figure showing an example of where normal diffusion occurs. MSD is plotted against time and the gradient of the slope is used to calculate the diffusion coefficient. Figure taken from reference 13.

The self-diffusivities can be calculated from the trajectory of MD simulations using the Einstein expression, Equation 2.18. These can be directly determined from MSD plots. An example of this is shown in Figure 2.5. When *t* is small, the motion of the gas molecule is considered ballistic and the logarithmic slope tends towards 2. After this, there is an intermediate region, where single file diffusion occurs; this has a logarithmic slope of 0.5. Normal diffusion is then seen when the logarithmic plot tends towards 1.0. Here, the MSD should increase in a linear fashion with respect to time.

2.2. Geometric surface area calculations

2.2.1. Types of surface area

When it comes to understanding the behaviour of porous materials, one key property is their surface area. For ordered porous solids, the structural model of this is generally determined using single crystal structure X-ray diffraction data, though if this is unavailable, then models based on other experimental measurements can be used; for example, X-ray powder diffraction. There are several definitions for surface area found within the literature, each with its own use. Generating a surface area can be conceptually considered as rolling a probe molecule along the interface of the atoms in the system. The most common two calculations are the Connolly surface area (CSA)[14] and the solvent surface area (SSA). The main difference between these is the point on the probe molecule from which the surface is drawn; the CSA is taken from the interface of the probe with the atoms within the system, whereas the SSA is taken from the geometrical centre of the probe.[15] The difference between the two surfaces is shown in Figure 2.6.

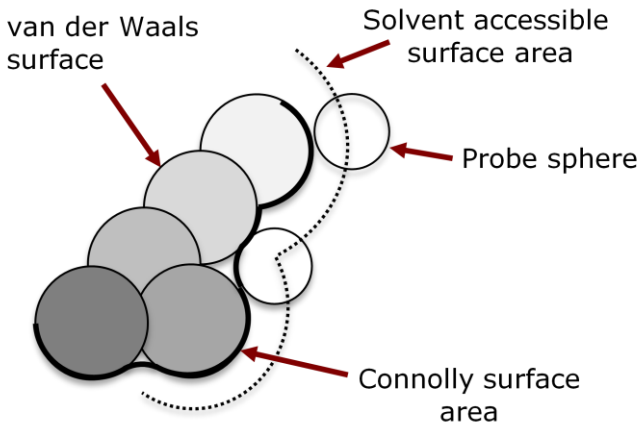


Figure 2.6. The van der Waals surface is measured from the edge of the van der Waals spheres. Conceptually, this is like "rolling" a probe along the van der Waals surface to generate two more surfaces; the solvent accessible surface and the Connolly surface. The solvent accessible surface is measured from the centre of the probe, while the Connolly surface is measured from the contact of the probe and the van der Waals surface.

2.2.2. Methods of calculations

There are two different methods that can be used to generate geometric surface areas from structural models; a Monte Carlo style approach, where a probe of a given radius is inserted at random into the structure, and a more systematic grid-based method. For both methods, the probe size remains constant and this

should correspond to the size of a specific guest; this is itself inserted into the system, and for each position a test is undertaken to see whether the probe is intersecting other atoms, or whether it is lying in free space. If the probe is lying in free space within the system, then the insertion is accepted and the value contributes towards the overall surface area of the pore structure; if it is intersecting with another atom, it is rejected and therefore not included. The chosen van der Waals radius of the host atoms is an important consideration and throughout this work these have remained constant; carbon = 1.70 Å, hydrogen = 1.09 Å and nitrogen = 1.55 Å.[16]

2.2.3. Grid method

Calculating the surface area using the grid-based method is the most systematic of techniques, because every point in the system is examined. Here a pre-determined grid interval is used to split the system into a 3D grid. For each point on the grid, the probe is placed and the selection criteria mentioned above is applied. This is then used to generate the surface area. Although this is more systematic than the Monte Carlo method, the computational expense can be greater. The quality of the resolution for the surface area is highly dependent on the size of the grid interval; a small value gives a finer resolution, whilst a higher value reduces the resolution. There is, of course, also an analogous trade off in terms of sampling for Monte Carlo methods.

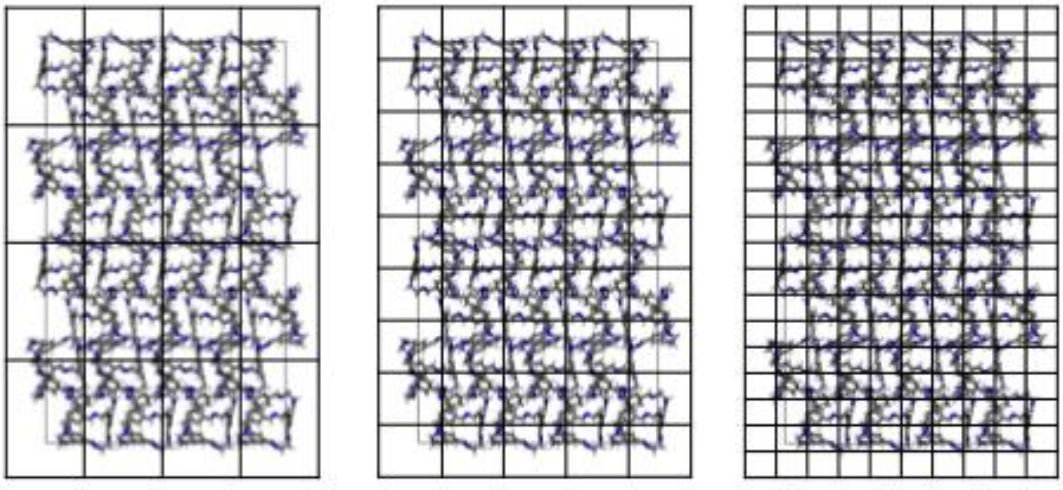


Figure 2.7. An illustration of how a grid method is applied to a system (**CC1a**). By decreasing the size of the grid interval (left to right), more space is sampled, thus describing the surface more accurately. However this increases the computational expense.

To give an example of the sensitivity to grid interval: if you reduce the grid interval from 0.4 Å to 0.1 Å, the calculation can take at least 3^4 times longer to

compute; this is the point where the value is converged within the error criteria.[17] Therefore a compromise between 'quality' and computational expense is required. The method is explained further in Figure 2.7. This method is used within Accelyrs' Materials Studio package.[18]

2.2.4. Monte Carlo method

This method is more commonly used, and is based on the Monte Carlo integration approach developed by Shrake and Rupley.[19] By sampling points of a sphere centred around each atom of the material, the accessible surface area can be established. This is achieved by summing up all the feasible spheres within the system. A sphere is deemed feasible if it does not overlap with another sampling sphere; each sphere has a total radius of the atom within the material plus the probe.

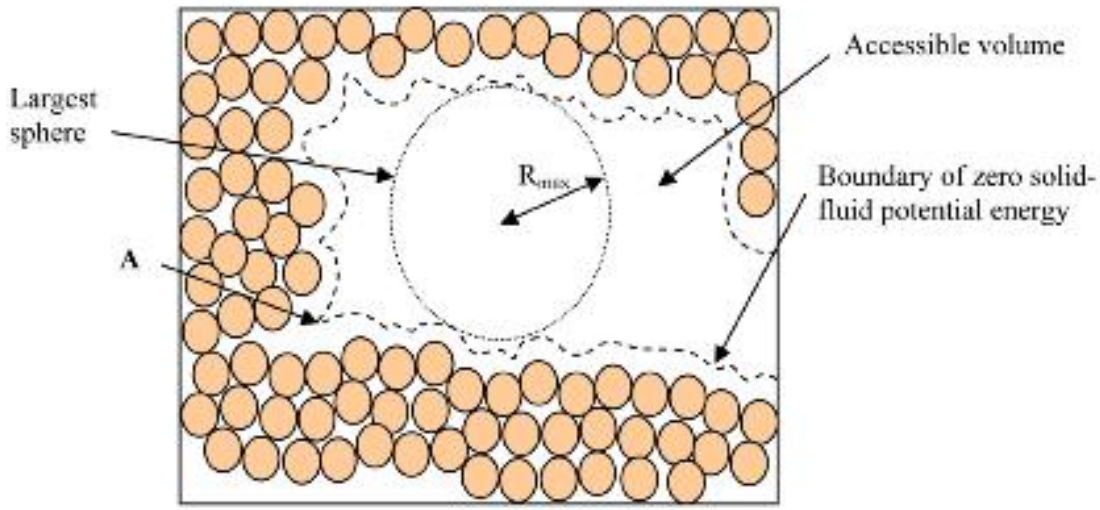


Figure 2.8. Illustration of a simulation box, where the dashed line represents the boundary at which the solid-fluid potentials energy equals zero; this is then labelled the accessible volume. Figure taken from Do *et al*.[20]

Finally, to calculate the surface, the fraction of feasible points is multiplied by the sampling sphere's surface area.[21] This approach, used by both Düren *et al*. and in the computer package Zeo++. It shows how channels that are too narrow to fit the probe are excluded from the calculated surface.[22] Do *et al*. used an alternative methodology, where the potential between the probe and its surrounding atoms was calculated. Only potentials that are negative or zero are deemed viable; this removes the chance of the probe overlapping the surrounding atoms.[23] This method is shown in Figure 2.8.

2.2.5. Comparison to BET model

Often, calculated geometrical surface areas are compared to experimental gas sorption values using either Langmuir or BET theory. Both of these theories assume close packed monolayer formation.[22] However, it is apparent that these comparisons can break down, especially for the Langmuir approach, as this only considers one monolayer of gas, which becomes invalid when the pores within the system are large enough to host several monolayers. Thus, calculated surface areas are often compared to BET isotherms; this allows for screening of various porous materials and benchmarking against materials in the literature.[24] Although this is common practice and does have its advantages, it not always strictly applicable, since a number of assumptions are made in the BET theory and these leas to inaccuracies. This is especially apparent for microporous systems with very open, high surface area structures. First, the BET method assumes that gas adsorption occurs in the formation of simultaneous monolayers and that there is no limit to how many layers can form at saturation pressure.[2] Any overlap between monolayer and multilayer adsorption would invalidate this.[24] In addition to this, it has been proposed that for high surface area porous materials, for example certain MOFs, a pore-filling mechanism is observed as opposed to multiple monolayer formation.[25] Another phenomenon that is overlooked is pore blockage; this is when gas is prevented from being adsorbed because gas has already filled the existing space.

There are a number of differences when directly comparing the BET method to simulated surface areas. First, the radius of the probe for BET methods differs to that used for calculating a surface area. For example the probe radius for nitrogen in the BET method corresponds to 2.3 Å (this corresponds to a cross-sectional area of 16.2 $Å^2$) whereas when simulating the surface area for a system, 1.82 Å is often used.[26] Also, the simulations assume slit pore topology; this assumes that the adsorbates adsorb in pores that are planar in geometry.[27] Finally, simulations model an ideal system, thus meaning no defects or distortions in the structure are considered.

2.2.6. Connectivity

Although modelling the geometric surface area for crystalline materials is important, and such calculations can be compared successfully to the BET model, there are a few minor limitations. One of these is that the connectivity of the system may be misrepresented. This is because the kinetic diameter of a gas is generally used to calculate the surface area. While this assumption is acceptable

for systems where the pore aperture is much larger than the gas, when the pores and gas are of a similar size, it is possible for pockets of void space to be incorrectly marked as inaccessible, giving an unrealistic picture of the pore topology.

To combat this, the connectivity of a system can also be determined separately. Instead of using the kinetic radius of a gas to generate the pore topology of a system, the gas is allowed to re-orientate to minimise its size. For example, for nitrogen the kinetic radius is 1.82 Å. However, as the gas molecule is linear in nature, the two atoms can be viewed end-on, thus giving it the effective radius of a single nitrogen atom, 1.55 Å. This value can therefore be used to generate the pore connectivity of a system, Figure 2.9.

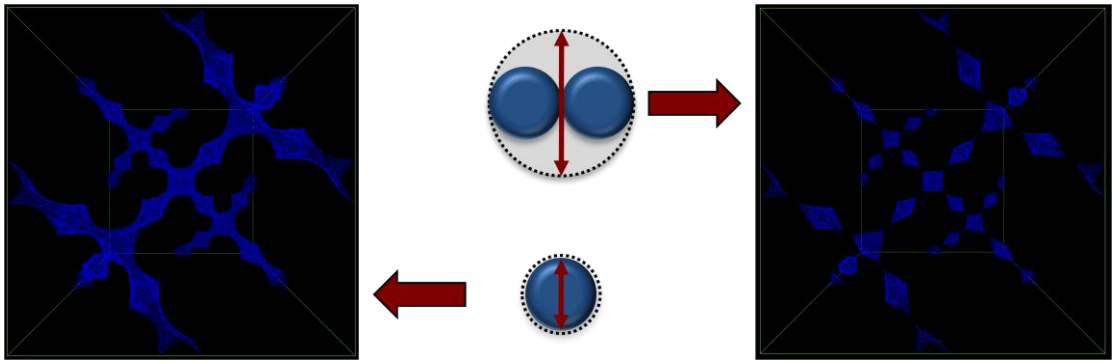


Figure 2.9. Shows a nitrogen molecule modelled using the kinetic radius (top) and end-on (bottom) and how this affects the pore connectivity for **CC3**. A connected pore topology is observed when using the end-on model (1.55 Å), left, while if the kinetic radius (1.82 Å) is used, the pore topology is disconnected.

2.2.7. Types of codes used to calculate surface area

One key aspect to consider when analysing porous materials is how to calculate their surface area. This can be done in a number of ways and the different styles of surface area have been previously discussed in Section 2.2.1. There are a number of programs that can be used to calculate the surface area of a given porous system. The first of these is a code developed by Düren that essentially calculates the accessible surface area by a simple Monte Carlo integration technique.[22,28] Another alternative is to use PoreBlazer, developed by Sarkisov.[29] This uses a similar methodology and can calculate both the accessible surface area and the pore size distribution. Both of these codes were developed using Fortran, though no visualisation was available for Düren's program; PoreBlazer though does output a visual representation of the pores.

Designed for the analysis of porous materials, Zeo++ was developed by Haranczyk and is based upon the Voronoi decomposition; this gives the

arrangement of the atoms in a periodic system and represents the void space as a graph.[30] This is a MC technique and the resulting Voronoi network can be used to calculate the largest included sphere (D_i), the largest free sphere (D_f), the largest included sphere along the free sphere path (D_{if}). The pore topology can then be determined. These quantities are shown schematically in Figure 2.10, as well as an example of a generated pore topology of **CC2**.[21]

Another major advantage of using Zeo^{++}, is that it can be used to either analyse a single structure, or to perform high-throughput analysis on a large database; it would be a good tool to use when screening a class of structures. In addition to this, adaption of the code through a collaboration with Haranczyk (Lawrence Berkeley National Laboratory) has led to a way to monitor dynamic connectivity in this thesis; viewing the topology of a dynamic structure opposed to the traditional dynamic structure. This is outlined later in this thesis.

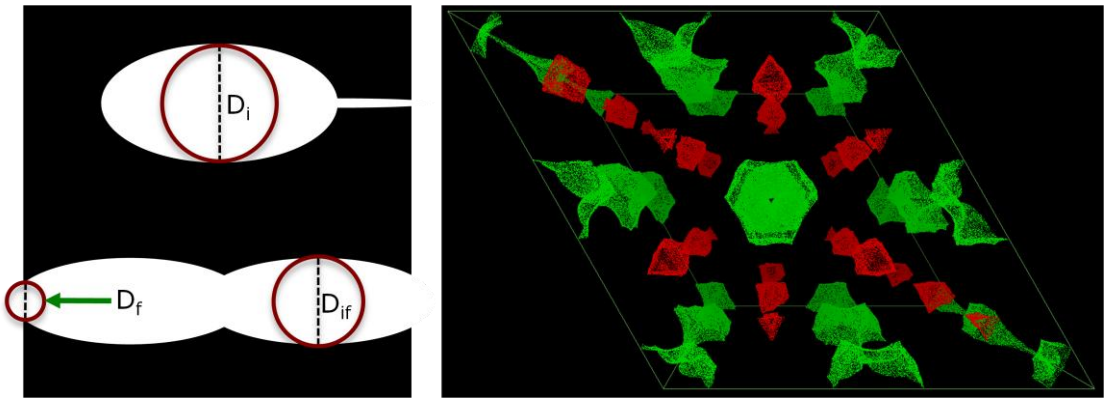


Figure 2.10. Illustration shows left, a scheme explaining the terminology used in Zeo++, and right, an example of a generated isosurface for a **CC2** (2220 isomer*).

2.3. Molecular simulations

2.3.1. Potential energy surfaces (PES)

Molecular simulations usually apply the Born-Oppenheimer approximation and this means that the electronic and nuclear motions can be separated. This is valid as the mass of an electron is far smaller and can therefore adjust rapidly to changes in the nuclear position. This means that the energy of a system, in its ground state, can be considered as an explicit function of its nuclear coordinates.[31] If these coordinates were to change, the energy of the system would consequently change; these changes can be isolated to something as simple as a bond stretch or as large as the movement of a cluster of atoms. The

* See Appendix A.1. for details on this.

magnitude of energy change would depend on what alteration had been made to the host system. It is therefore possible to track the energy expense of specific parameters systematically. In a force field (FF), a parameter is a single component within the system; for example a bond stretch or angular rotation. This is achieved by altering a single parameter by a set amount, constraining it and subsequently allowing the rest of the parameters within the system to find their ground states; the energy difference between the updated system and the ground state corresponds to the energy required to move the parameter by the set amount. This can then be plotted as energy versus parameter change and shown graphically to measure how significant that parameter is. If the energy change is large, then the alteration to the structure is significant; if it is only small, then the change is relatively insignificant. These energy changes can be classed as movements on a multidimensional surface; a potential energy surface (PES).[2]

2.3.2. Energy minimisation

The potential energy of a system is complicated because it involves the sum of the multidimensional functions for all of the atom coordinates. For a system containing N atoms, the energy is directly related to a function of 3N; 3N cartesian coordinates or 6 internal coordinates. This makes it difficult to visualise this function for anything other than the simplest of systems; generally a few atoms, for example a hydrogen molecule. It is, however, possible to focus on specific parts of the energy surface, for instance a bond rotation; the energy here can be described as a function of two variables, and from this contour diagrams can be visualised.[2]

Many problems found in computational chemistry can be considered as a minimisation of a multi-dimensional function.[17] This style of minimisation is the identification of the stationary points of a function; a stationary point is when the first derivative is zero. The most desired of these are generally the minimum points, when the second derivatives are positive. Another type of stationary point is the maximum; this is where the second derivatives are all negative. Finally, there are also saddle points; here the second derivative is negative in one, but positive in all other directions.[17] These are illustrated in Figure 2.11. These minimum points correspond to stable conformers of the structures minimised – the saddle points are transition states, and the global minimum is the ground state for the structure.

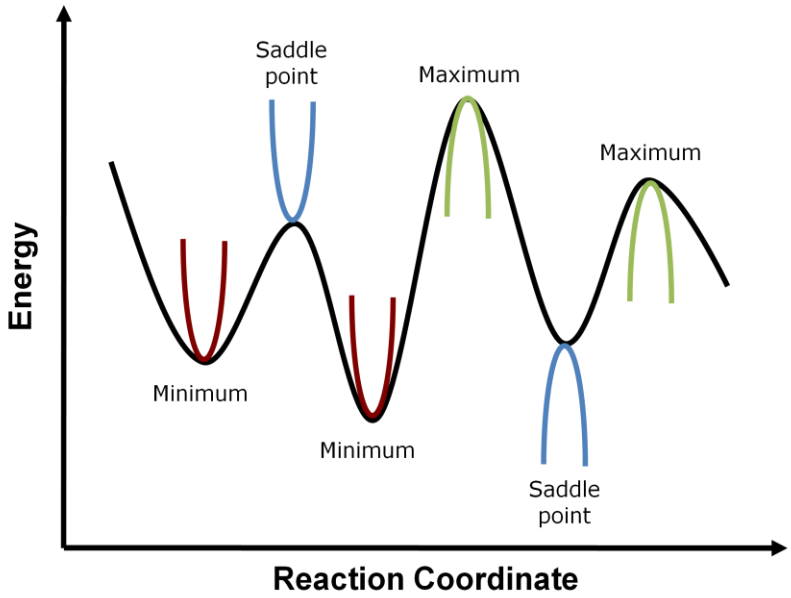


Figure 2.11. Illustration of a multidimensional PES, showing maxima, minima and saddle points.

The multidimensional function may contain a number of stationary points; these all help identify the most and least favourable arrangements of the atoms. The global minimum of the function is the lowest value on the entire PES – all other stationary points where the second derivatives are all positive are known as the local minima. As minimisation techniques locate the nearest stationary point, it does not necessarily mean that the minimum obtained is the true global minima of the system but could instead be a local minimum.

Different minimisation algorithms are available, with the most common being either first-derivative or second-derivative techniques; suitability is dependent on the problem involved and will be chosen for both its efficiency and accuracy.[17] The first derivative gives an indication to the direction of the minima and the steepness of the slope at this given point. The second derivative provides additional information about the curvature of the surface. Two examples of first-derivative minimisation algorithms are the conjugate gradient method and the steepest descent method. For each minimisation step of the conjugate gradient method, the gradient is calculated; the next minimisation step is then given a direction conjugate to this. This is cycled until the convergence criteria are met. This is good for large scale systems because it has a higher level of accuracy when compared to the steepest descent method; it is also computationally cheaper. The steepest descent minimisation method differs to the conjugate gradient method. The gradient of the first step is calculated and then any subsequent optimisation steps are along the line of steepest descent. This

method is often used as a rough optimisation process when the initial configuration is located far from a minimum. Both of these methods are cheaper than their second-derivative counterparts; however, they do not give any information about the curvature of the surface and therefore can often be inefficient and lower in accuracy. One second-derivatives method is the Newton-Raphson method.[32] This is the simplest second-derivative method available and calculates the inverse Hessian matrix – this is a matrix containing the second-derivatives. Although computationally more accurate, the inversion and storage of the Hessian matrix is computationally expensive and demanding on memory. As this has to be calculated for each optimisation step, the Newton-Raphson method is not viable for larger systems. One way to get round this is to use a Quasi-Newton method. These were developed so that the full Hessian matrix does not need to be calculated for each iteration, thus drastically reducing the computational expense.[17]

During this work, the Smart minimiser within Materials Studio 5.0 was used. This combines the steepest descent, conjugate gradient and Newton-Raphson method in a cascade. The robust but less accurate first-derivative algorithms are used first to get close to the minimum and then the more accurate second-derivate Newton-Raphson method is used close to the end of the run to increase the accuracy.

2.3.3. Interatomic potential calculations

One potential problem with using large systems when the motion of the electrons within the system are considered is that the computational expense to run anything on a meaningful timescale can be enormous. Hence, the use of interatomic potential methods are more desirable, as the Born-Oppenheimer approximation[2] is used so that each atom is modelled as a classical particle; the motion of the electrons is ignored. This states that as the nuclear motion is much slower than the electronic motion and consequently only the nuclear position needs to be established, as the electrons will instantaneously accommodate the nuclear movement. By removing the complication of the electronic motions, simulations of larger systems becomes more viable and longer time scales can be considered. Interatomic potentials work by dividing up the potential energy of a system, E_{total}, into its individual components. A force field (FF) divides the potential energy up into a number of different components. These components generate an energetic penalty when they deviate from their ground state. This could be the overstretching of a bond, the over rotation of an angle, the extra

rotation of a specific bond, or, potentially, the interaction of non-bonded atoms. One positive feature about using a FF is that the internal configurations within the system can be monitored (for example a bond length) and the alteration to the parameter associated with this can then be quantified directly. This makes it easier to grasp how modifications in the FF parameters affect the performance, while assisting the parameterisation process. The functional form for the sum total of the energy (E_{total}) for the FF can be expressed as:

$$E_{Total} = E_{stretch} + E_{bend} + E_{torsion} + E_{electrostatic} + E_{VDW} \quad \text{(Eq. 2.22)}$$

This can be separated into the intramolecular (bonded) and intermolecular (non-bonded) potentials:

$$E_{Total} = E_{bonded} + E_{non-bonded} \quad \text{(Eq. 2.23)}$$

The specific energy contributions within these are:

$$E_{bonded} = E_{stretch} + E_{bend} + E_{torsion} \quad \text{(Eq. 2.24)}$$

$$E_{non-bonded} = E_{electrostatic} + E_{VDW} \quad \text{(Eq. 2.25)}$$

Where E_{bonded} is the sum of the two-body stretching potentials ($E_{stretch}$), the three-body bending potentials ($E_{bending}$) and the four-body torsional potentials ($E_{torsion}$), found within a system – these are the intramolecular contributions of the system. $E_{non\text{-}bonded}$ is the sum of the intermolecular potentials – the electrostatic interactions ($E_{electrostatic}$) and the van der Waals interactions (E_{VDW}). These specific energy contributions are described in Figure 2.12.

One goal of a FF is to reproduce the structural properties of a system, though it is possible to use it to predict other properties, for example molecular spectra. One important feature of a FF is its transferability to other similar systems. The FF becomes much more desirable if it can accurately describe a number of similar systems, rather than having to re-parameterise the parameters for every single system, which would be time consuming. The only issue is that if a FF becomes too generic, the accuracy is impacted, and therefore there needs to be a balance between transferability and accuracy. A key point when using a FF, is that it is empirical in nature and consequently this means that there is no 'correct' form. The choice of functional is generally down to their computational efficiency, as well as how precisely they model the date their attempting to describe.

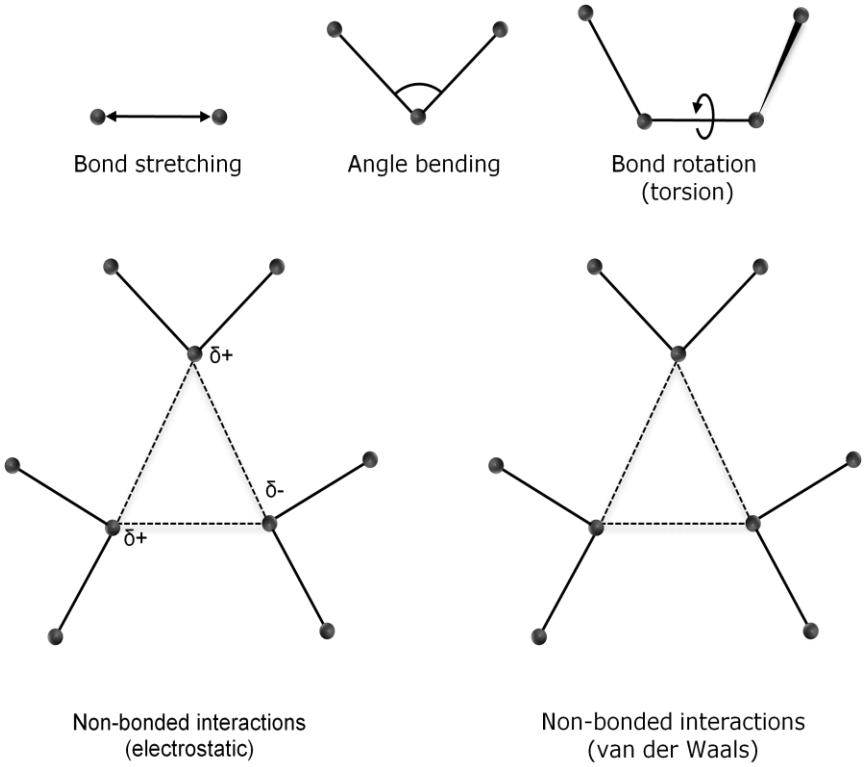


Figure 2.12. Schematic representation of the different contributions to a molecular mechanical FF; bond stretching, angle bending, torsional terms and non-bonded interactions. Figure adapted from reference 2.

2.3.4. FF atom types

So that each atom type within a system can be accurately and appropriately described, they are assigned an individual FF atom type. This gives an indication into the nature and properties of the atom. A range of properties can be used to define which atom type to use for which atom and these could include:

- The element type.
- Types of bonding involved; this could be single, double, triple, resonant, and so on.
- Number of atoms that are bonded to specific atom.
- The types of atoms that the given atom is bonded to.
- The hybridisation.
- The formal charge on the atom.

Since the typing of the atoms is only based on their local environment, there is a possibility that many atoms within the same system could have the same FF atom type, especially if there is symmetry within the system. Additionally, sets of two, three, and four atom strings allow specific parameter sets for bonds, angles and dihedrals, to be applied. PES can then be generated for these parameters, using the appropriate level of theory (discussed later), and subsequently fitted

using the equations found within the FF to simulate the behaviour of these atoms. Therefore, accurate FF atom typing is imperative if the FF is to be accurate.

2.3.5. Bond stretching

The first term in Equation 2.22, $E_{stretch}$ represents the bond stretching contribution in the FF. There are a number of options when choosing the functional form used to represent the bond stretching. One of these in the Morse potential:[2]

$$E_{stretching} = D_e\{1-\exp[-a(l-l_0)]\} \quad \text{(Eq. 2.26)}$$

$$a = \omega\sqrt{\mu/2D_e} \quad \text{(Eq. 2.27)}$$

Where D_e is the depth of the potential well, μ is the reduced mass, ω is the angular frequency of the bond and l_0 is the reference value of the bond; ω can be related to the stretching constant in the following manner:

$$\omega = \sqrt{k/\mu} \quad \text{(Eq. 2.28)}$$

The relatively complex nature of the Morse potential gives it the ability to describe a wide variation in the parameter – from a small to large vibration around the equilibrium point; this enables it to model strong equilibrium behaviour whilst at the same time it has the ability to model dissociation. The main drawback with using the Morse potential is that it is computationally expensive and requires three constants for every bond. This means that it is rarely used in FFs.

The forces between two bonded atoms are very large and therefore a large amount of energy is required to make these deviate from their equilibrium position. As bonds tend not to deviate too much from their ground state, a simpler, more harmonic potential like Hooke's law can be used:

$$E_{stretching} = \frac{k}{2}(l-l_0)^2 \quad \text{(Eq. 2.29)}$$

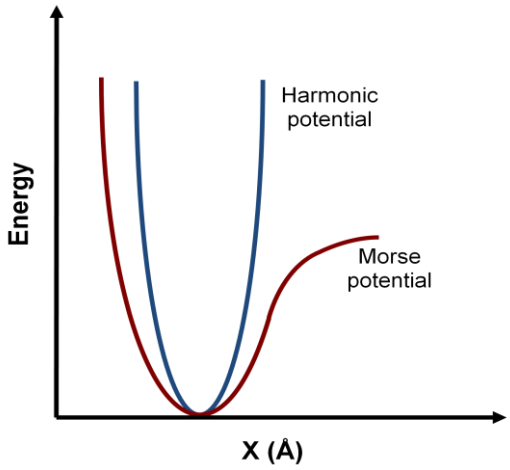


Figure 2.13. Scheme showing the difference between a harmonic potential and a Morse potential.

This simply states that the energy of $E_{stretch}$ is a function of the square distance away from l_0; the reference bond length. This reference bond length is the value of the bond when all other parameters within the FF have been set to zero. This is different to the equilibrium bond length; this is the value of the bond when the whole system is found in its ground state. These values often deviate slightly and this is down to the intricate interplay between all the parameters; l_0 for one parameter may have to shift away from the reference value in order to compensate for another parameters. Often the system modifies all the l_0, so that the energetic minimum can be obtained. It is important to note that l_0 should not be a singular number as vibrational motion is present for 'real' molecules. Although Hooke's law gives a reasonable accuracy for the well depth around the equilibrium point of the bond stretching potential, the harmonic nature tends to lead to inaccuracies when the energy moves away from this minimum point. To combat this, a Morse potential curve can be mimicked by the addition of quadratic, cubic, quartic, and higher level terms. This leads to the expression:

$$E_{stretching} = \frac{k}{2}\left(l - l_0\right)^2 \left[1 - k'\left(l - l_0\right) - k''\left(l - l_0\right)^2 - k'''\left(l - l_0\right)^3 \ldots\right] \qquad \text{(Eq. 2.30)}$$

If only quadratic and cubic terms are included in the new potential, the potential now has a maximum and this can lead to unrealistic bond lengths. The difference between a harmonic potential and a Morse potential is shown in Figure 2.13. An example of such a FF is MM2; this is valid if the structures are close to equilibrium, but lead to inaccuracies if the 'true' potential well is missed. To remove this issue, a quartic term can also be included. This eliminates the inversion problem seen with limiting the potential to a quadratic and cubic term

and an example FF is MM3.[33] This produces a curve much more representative of a Morse curve.

In this work, a variation of this, the quartic potential, is used. Here, r and r_0 are the current and reference bond length and k_2, k_3, and k_4 are the variables for the quadratic, cubic, and quartic terms:

$$E_{stretching} = k_2\left(r - r_0\right)^2 + k_3\left(r - r_0\right)^3 + k_4\left(r - r_0\right)^4 \quad \text{(Eq. 2.31)}$$

2.3.6. Angle bending

The description of the angles away from their reference values also uses Hooke's law, so that:

$$E_{bending} = \frac{k}{2}\left(\theta - \theta_o\right)^2 \quad \text{(Eq. 2.32)}$$

Where θ and θ_0 the current and reference bond angles respectively. A lot less energy is typically required to move an angle away from the equilibrium, as it is harder to stretch or compress a bond; therefore the force constants will be smaller in value. This means that using a simple harmonic potential will become inaccurate more easily when defining an angular potential and therefore higher-order terms need to be included. Therefore, like with the bond stretching potentials, a Morse potential curve can be mimicked by the addition of quadratic, cubic, quartic and higher level terms:

$$E_{bending} = \frac{k}{2}\left(\theta - \theta_0\right)^2\left[1 - k'\left(\theta - \theta_0\right) - k''\left(\theta - \theta_0\right)^2 - k'''\left(\theta - \theta_0\right)^3 \ldots\right] \quad \text{(Eq. 2.33)}$$

Like before, within this work quartic potentials have been used to describe all the angles within the system. Here, θ and θ_0 the current and reference bond angles and k_2, k_3, and k_4 are the variables for the quadratic, cubic, and quartic terms:

$$E_{bending} = k_2\left(\theta - \theta_0\right)^2 + k_3\left(\theta - \theta_0\right)^3 + k_4\left(\theta - \theta_0\right)^4 \quad \text{(Eq. 2.34)}$$

2.3.7. Torsional terms

Both $E_{stretch}$ and $E_{bending}$ are sometimes referred to as 'hard' degrees of freedom; this is because a relatively large amount of energy is required to bring about a movement away from the reference values. $E_{stretching}$ is 'harder' then $E_{bending}$, as more energy is required to deform a bonds length, but $E_{bending}$ is 'harder' than $E_{torsion}$. This means the dynamic nature of the system is generally dictated by the

contribution and complex interplay of the torsional and non-bonding characteristics of the system. $E_{torsion}$ is different to the other intramolecular terms, as there is potential for multiple maxima and minima. They follow a cosine expression and multiple variations are available. The one used within this work is a three-term cosine potential and this has the expression:

$$E_{torsion} = \sum_{3}^{n=1} \frac{V_n}{2} \left[1 - \cos\left(n\phi - \partial_n \right) \right] \quad \text{(Eq. 2.35)}$$

Where V_1, V_2, and V_3 are the barrier heights and Φ is the current torsion angle; Φ_{01}, Φ_{02}, and Φ_{03} are the reference torsion angles in degrees. The reference angles are usually 0° or 180°.

2.3.8. Intermolecular potentials

The van der Waals forces contribute towards the long-range dispersive forces in combination with the short-range repulsions seen between atoms. The long-range dispersive forces come about by the generation of instantaneous dipoles located on the atoms; these have a knock on effect on atoms in close proximity, inducing a change in their own dipole moment. The short-range repulsion between the atoms are generated using the Pauli exclusion principle; if two atoms are close enough that their electronic orbital's overlap, the Pauli exclusion principle prohibits an electron from being placed in both orbitals, as they cannot occupy the same region of space.

The van der Waal forces can be modelled using a variety of simple potentials; the main criterion is that it incorporates both the long-range and short-range contributions. The most commonly used potential is the Lennard-Jones potential. One example of this is the 12-6 Lennard-Jones interaction – this is used for a wide range of organic systems.

$$E_{VDW} = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \quad \text{(Eq. 2.36)}$$

Here σ is the collision diameter and ε is the potential energy well depth; the collision diameter is the ideal atom-atom separation. In this work a variation of the above equation is used; this is known as the 9-6 Lennard-Jones potential and has the form:

$$E_{VDW} = \varepsilon \left(2\left(\frac{\sigma}{r} \right)^{9} - 3\left(\frac{\sigma}{r} \right)^{6} \right) \quad \text{(Eq. 2.37)}$$

It is possible to scale the dispersion characteristics within the FF. This is achieved by altering the section of the equation that controls the dispersion; the r_6 part. Therefore the r^6 and r^9 parts can be split up, with respect to the potential:

$$E_{VDW} = \frac{A}{r^9} - \frac{C}{r^6}$$

(Eq. 2.38)

Here, Both '*A*' and '*C*' are constants, with '*C*' being responsible for the dispersion. These constants are defined as:

$$A = 2\varepsilon\sigma^9 \qquad \text{and} \qquad C = 3\varepsilon\sigma^6 \qquad \text{(Eq. 2.39)}$$

or,

$$\sigma = \left(\frac{3A}{2C}\right)^{\frac{1}{3}} \qquad \text{and} \qquad \varepsilon = \frac{4C^3}{27A^2} \qquad \text{(Eq. 2.40)}$$

This means that to scale the dispersion, the '*C*' parameter needs to be altered. This allows the FF to be adapted to any given system, if required. The style of equation used to describe the dispersion means that it is long range in nature and therefore cutoffs are often introduced to improve efficiency. It is important that an appropriate cutoff is chosen; otherwise important intermolecular parameters could be removed and the system not well represented.

The other intermolecular contribution that needs to be considered is the $E_{electrostatic}$. The atoms act as point charges and $E_{electrostatic}$ is calculated using:

$$E_{electrostatic} = \sum_i \sum_j \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{(Eq. 2.41)}$$

Where q_i and q_j are the partial charges of atoms '*i*' and '*j*' and r_{ij} is the distance between the two atoms. This is Coulomb's law and is difficult to calculate, as there is an inverse dependence on r_{ij}. Their long-range nature makes them challenging to calculate. One major problem is that the Coulomb interactions only become conditionally convergent when the series of positive and negative terms become divergent.[17] To simplify this, the electrostatics can be truncated using a specific cutoff. This can be complicated though as there is the sizeable variation in the Coulomb terms at shorter atomic distances. One method to calculate $E_{electrostatic}$ is to use the Ewald summation method.[34]

2.3.9. Methods for fitting force fields

FFs consist of a number of intramolecular and intermolecular parameters that aim to describe the motion of a system or set of systems. The advantage of using mathematical potentials to describe the interactions within a system, is that the computational expense is greatly reduced, especially as the electronic density is not explicitly considered. The disadvantage is that the accuracy of the FF is highly dependent on the data that the parameters are fitted to. There are a number of alternatives:

1. Experimental data, such as crystal structures and vibrational frequencies, can be used to find both the ground state and oscillation of a system. The crystal structure can be used to fit the intermolecular dispersion parameters, whereas the intramolecular parameters can be modelled on the bonds, angles and dihedrals found within the system.

2. Spectroscopic data, such as NMR, microwave spectroscopy, and IR data can be used.

3. The geometries and relative conformational energies of key molecules are often inserted into a FF, though geometric data for metal coordination compounds is scarce.

4. Some FFs are fitted to reproduce thermodynamic properties, for example the OPLS FF parameters were developed and tested using the conformational energetics and properties of organic liquids.[35]

5. If experimental data is unavailable, quantum mechanic calculations can be used to provide the data. The parameters can be fitted to higher level *ab initio* data. This often involves the use of a fragment within a system, as modelling the whole system would be computationally too expensive. PESs are generated for all the internal degrees of freedom within the fragment and this data is then used to fit to a harmonic potential; this potential is subsequently used within the FF.

6. A combination of both high level calculations and experimental data can be used. FFs can be fitted against high level calculations are subsequently tested against experimental data for validation.

First-generation FFs, including AMBER,[36] CHARMM,[37] CVFF,[32,38] and MM2,[39] were developed using various fitting procedures with the aim of reproducing reference data from *ab initio* calculations and/or available X-ray crystal data—for example,

single crystal structures, calometric data, or vibrational frequencies.[40] The use of *ab initio* calculations to supplement X-ray crystal data has the advantage of directly providing data points on the PES of the system against which the FF is to be matched. Additionally, for large molecules, the fitting process may only need to consider small fragments of the target system that represent the key functional groups.[41-43]

Within this work, a combination of both high level calculations and experimental data has been used to fit a cage specific FF (CSFF). Here, the intramolecular potentials are modelled on PES generated for the internal degrees on freedom of a common cage fragment, whilst the intramolecular parameters were modelled on crystal structure data.

2.3.10. Periodic boundary conditions (PBC)

The porous organic cage systems investigated here are crystalline in nature and are therefore periodic; hence, for both MC and MD simulations, periodic boundaries need to be taken into account. To accurately reproduce the bulk properties of these systems, a huge number of atoms would be required; to generate such systems is computationally prohbitive. Therefore periodic boundary conditions (PBCs) are applied. This allows the system to be represented as a smaller cell that in turn experiences the same kind of forces that would be experienced if the system was much larger. This is achieved by creating a periodic image of the smaller system. This is subsequently transposed infinitely as a lattice in all directions. As a result there are no interfaces, as each atom can interact with its own image in the surrounding repeated systems.

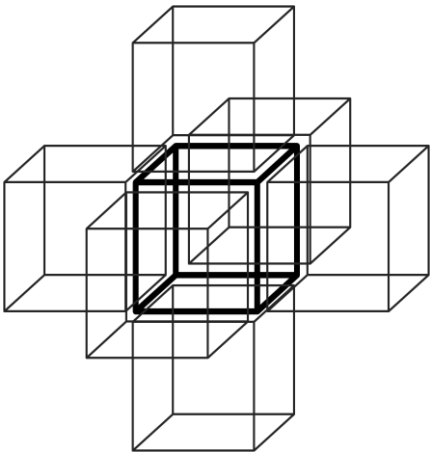


Figure 2.14. Scheme showing how periodic boundaries repeat the initial unit cell infinitely in every direction.

An example of this can be seen in Figure 2.14. As the number of particles within the cell remains constant, all that needs to be stored is their positions within the cell. One important thing to consider is the non-bonding interactions within the periodic system; as PBCs generate an infinite system, a cutoff distance is required so that the calculation of the energy remains tractable. Another way to reduce the amount of non-bonded interactions is to apply a technique known as the minimum image convention. Here, each atom only takes into account the non-bonded interaction with every other atom once; there is no discrimination between whether the other atom is located within the same or a neighbouring system. There are limitations to this approximation. By definition, the model cannot capture phenomena on length scales longer than the simulation cell. Also, the models are 'perfect', and features such as crystal grain boundaries and defects will not be accounted for.

2.4. Electronic structure calculations

It is possible to use other techniques to simulate chemical systems; instead of using more classical techniques like FF, quantum mechanics can be used. This helps to portray the electron distribution within each system in a level of detail not possible in classical calculations – this though becomes computationally more expensive. This is based on the Schrödinger equation, which has the expression:

$$\hat{H}\psi = E\psi \qquad \text{(Eq. 2.42)}$$

Where $\hat{H}$ is the Hamiltonian operator, E the total energy of the system, and ψ the wavefunction. It is only possible to exactly solve this equation for the simplest of systems and this is because of the 3N dependency upon the number of atoms. This is brought about by the interactions between the electrons in the system. This means that to calculate the energy of a large system, approximations need to be made; one of these is the Born-Oppenheimer approximation, as defined above. This in turn permits the electronic part of the equation to be solved using the nuclear positions of the atoms as parameters, generating PES for these which can be used to define the nuclear motion of the system. The computational expense is dictated by solving the electronic part of the Schrödinger equation for a system.

As the dynamics of many-electron systems are complex, it is possible to conceptualise the system such that the motion of one electron is considered independent to the dynamics of the other electrons within the system; this is known as an independent-particle model. Using this model results in an

approximation for the interaction between particles, achieved by taking all the interactions within the system and averaging them; this is known as Hartree-Fock theory.

2.4.1. Hartree-Fock (HF) methods

HF theory is the foundation of molecular orbital theory and expands the Schrödinger equation so that the wavefunction is equivalent to a single determinant; every electron is portrayed using an orbital and the wavefunction is described as a product of these orbitals. The electron correlation effects are ignored. The nuclei in these are considered to be stationary and the electrons are found in their ground state. These orbitals are arranged in a Slater determinant, so that the function is asymmetric. The orbitals shape describes a region of space where it is probable to find an electron; here the average repulsion to all other electrons is included in combination with the attraction to the nuclei. As the description of an electron is dependent on the surrounding orbitals, it must be solved iteratively. A self-consistent field (SCF) procedure is then used to solve the Slater determinant. The HF method is generally used to calculate the energy of smaller molecules; as the electron correlation is ignored, larger systems are often inaccurate. This model acts like a branching point to other levels of theory; this is illustrated in Figure 2.15. The options are the addition of new approximations, which brings forth semi-empirical methods, or the addition of more determinants, which helps to converge the Schrödinger equation to an exact solution. If the energy correlation effects are deemed important, post HF methods can be used to increase the accuracy of the results.

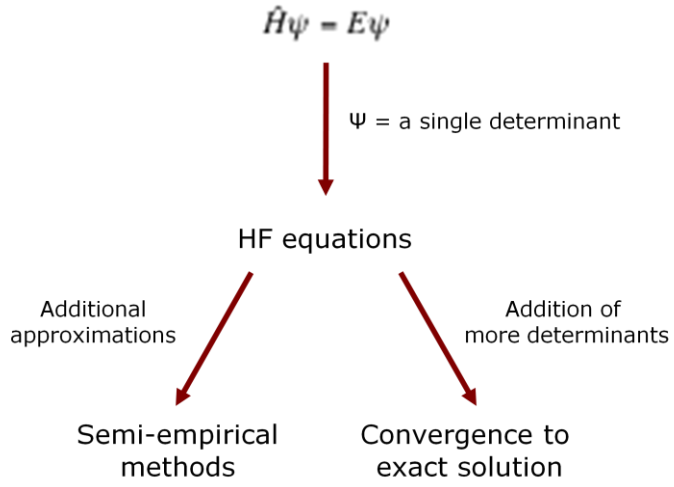


Figure 2.15. Using the HF model as a starting point, it is possible to either add more approximations to increase the speed of the calculation or to add more determinants to increase the accuracy.[17]

Semi-empirical data is less accurate than HF methods, as it ignores all the integrals involving two or more nuclei when constructing the Fock matrix. Success with semi-empirical methods is found when the ignored integrals are turned into parameters, which are subsequently fitted to experimental data, such as molecular energies and geometries. The addition of more determinants may give an exact solution for the Schrödinger equation, but the computational expense is greatly increased.

2.4.2. Density functional theory

HF methods focus on multi-electron wavefunctions, expressed via the Slater determinant; this is made up of a set of single-electron wavefunctions. Density functional theory (DFT) also uses these single-electron functions, though in a different manner. Unlike HF methods, DFT only calculates the total electronic energy in combination with the overall electronic density distribution; DFT assumes that there is a relationship between these. Thomas and Fermi originally touted this notion in the late 1920's, and Hohenberg and Kohn took up this work in the late 1960's.[2] Here they managed to illustrate that the electron density can be used to define the ground state of a system and that there was a mathematical basis for this theory.[2] The Hohenberg-Kohn theorem showed that if a system has a non-degenerate ground state, the electronic density of a system, $\rho(r)$, can be used to determine the external potential, V_{ext}. The total energy can be expressed as:

$$E[\rho(r)] = F[\rho(r)] + \int V_{ext}(r)\rho(r)dr \qquad \text{(Eq. 2.43)}$$

Here, $F[\rho(r)]$ is the universal functional and depends on solely upon the electron density. The universal functional can be defined as follows:

$$F[\rho(r)] = T[\rho(r)] + J[\rho(r)] + E_{XC}[\rho(r)] \qquad \text{(Eq. 2.44)}$$

Where $T[\rho(r)]$ is the kinetic energy, $J[\rho(r)]$ is the classical electron-electron interaction energy and $E_{XC}[\rho(r)]$ non-classical contributions from exchange and correlation. Out of the three terms only $J[\rho(r)]$, sometimes known as the Coulomb contribution, can be calculated explicitly. This is achieved by:

$$J[\rho(r)] = \frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r'-r|}drdr' \qquad \text{(Eq. 2.45)}$$

Considering $T[\rho(r)]$ first, calculating this mathematically is difficult, as the relationship between the electron density and the kinetic energy is not well

understood. To work round this the Kohn-Sham approach can be used. Instead of trying to explain the kinetic energy of electrons within an interacting system, the kinetic energy of electrons within a non-interacting system was instead calculated; for the two systems the electron density remained constant.[2] As the kinetic energy for electrons in a non-interacting system can be calculated using the electronic density, any overlooked energy associated with the interaction between the electrons can be added on using another term, the exchange-correlation energy, E_{XC}. This leads to the following equation:

$$\rho(r) = \sum_{i=1}^{N} \left|\psi_i(r)\right|^2 \qquad \text{(Eq. 2.46)}$$

Where N is the total number of occupied orbitals and ψ_i is the wavefunction of the electronic state; this equation assumes that the system is spin-independent. ψ_i can then be solved using the Kohn-Sham equation:

$$\left[\left(\frac{\mathrm{h}^2}{2m}\right)\nabla^2 + V_{KS}(r)\right]\psi_i(r) = \varepsilon_i\psi_i(r) \qquad \text{(Eq. 2.47)}$$

Here ε_i is the Kohn-Sham energy and V_{KS} is the external potential of the non-interacting system. V_{KS} can be calculated by:

$$V_{KS}(r) = V_{ext}(r) + \int \frac{\rho(r')}{|r - r'|}dr' + V_{XC}(r)$$

(Eq. 2.48)

Here, $E_{KS}[\rho(r)]$, the Kohn-Sham functional, was minimised with respect to $\rho(r)$. V_{XC} is the potential resultant obtained from the exchange-correlation energy and can be defined as:

$$V_{XC}(r) = \frac{\partial E_{XC}[\rho(r)]}{\partial\rho(r)} \qquad \text{(Eq. 2.49)}$$

By substituting in the wavefunction the kinetic energy can be calculated:

$$T_{KS} = -\frac{\mathrm{h}^2}{m}\sum_{i=1}^{N}\left\langle\psi_i\middle|\nabla^2\middle|\psi_i\right\rangle \qquad \text{(Eq. 2.50)}$$

Where T_{KS} is the kinetic energy from the Kohn-Sham orbitals. With this, it is now possible to write an expression for the Kohn-Sham functional, where only the exchange-correlation is unknown; this has the form:

$$E_{KS}[\rho(r)] = T_{KS} + \int \rho(r) V_{ext}(r) dr + \frac{1}{2} \iint \frac{\rho(r)\rho(r')}{|r_i - r_j|} dr dr' + E_{XC}[\rho(r)] \qquad \text{(Eq. 2.51)}$$

Calculating the DFT energy is an iterative process and this is shown as a flowchart in Figure 2.16.

2.4.3. Exchange and correlation functionals

As E_{XC} is unknown for all systems, except those exhibiting a homogeneous electron gas,[2] an approximation is required. This approximation describes the exchange and correlation energy, as well as the kinetic energy difference between the interacting and non-interacting system. There are two main ways to describe these approximations; the Local Density Approximation (LDA) and the Generalised Gradient Approximation (GGA). Considering the LDA first, here the potential depends solely on the electron density of the system at a given point – the derivatives are ignored thus localising this approximation. Originally proposed by Kohn and Sham,[44] it regards the inhomogeneity of the electronic system as homogeneous. As the electronic potential for homogeneous electron gas is known, this can be substituted in. This approach works well for both solids and molecules and its performance is widely documented.[44] One of the major drawbacks is that using LDA can lead to a poor description of weakly bonded inhomogeneous systems. The main reason for this is the diverse electron distribution within the system being treated as uniform.[2] This can lead to overestimation of both phonon frequencies and cohesive energies, whereas underestimation can be seen in the lattice parameters and band gaps of the poorly represented systems. Hydrogen bonding is also poorly described.

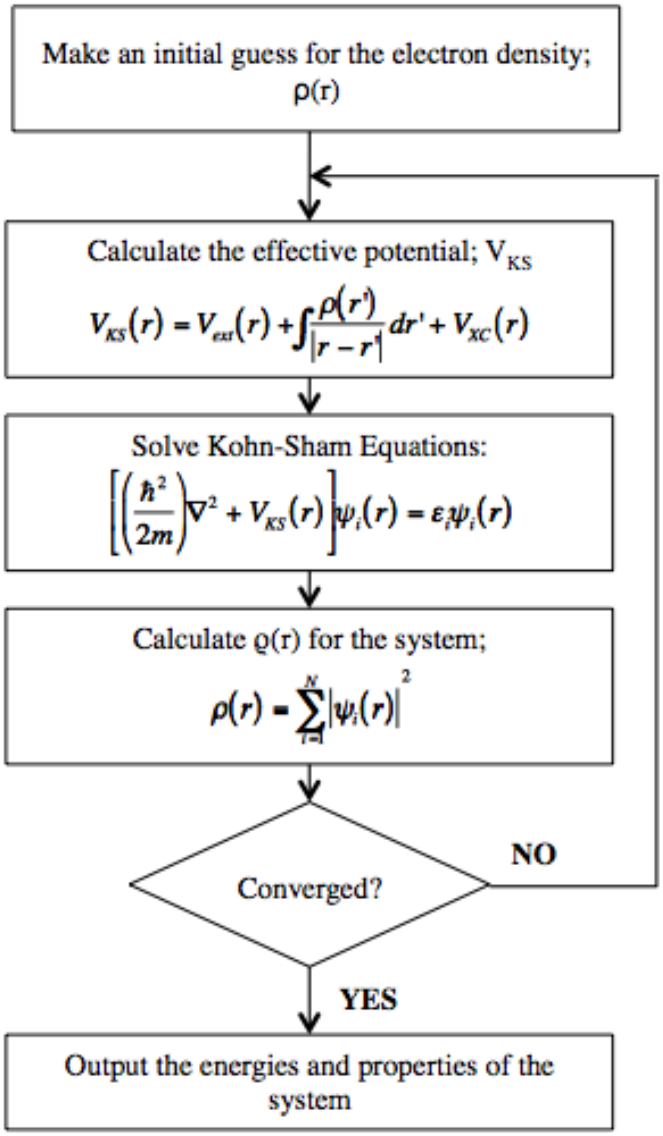


Figure 2.16. Flowchart showing iterative process involved with DFT.

Due to these limitations, it is important to improve upon this method. This is achieved by also describing the gradient and derivatives of the density; this removes the assumption that the system is homogeneous. One example of this is the GGA and this includes up to the second order gradient expansion. This method has been improved upon greatly over the past few decades.[2] The GGA often offers improved results when compared to LDA methods, with the energetics and bond angles and lengths of the systems being better described.[2] One example of a GGA is BLYP; here the Becke exchange functional[45] is combined with the correlation functional developed by Lee, Yang and Parr.[46] Another example is the PBE functional. This was developed by Perdew, Burke and Ernzerhof in 1996 and although it is based on LDA it attempts to address inhomogeneity by including most of the energetically important features.[47] A

different approach is to use a hybrid functional; these combine both DFT and HF methods. Here the exchange correlation energy incorporates a fraction of the HF exchange.[48] The rationale behind this is that whilst DFT may underestimate parameters like bond lengths, while HF methods tend to overestimate them; therefore a combination of the two would provide a more accurate result. This can be expressed as:

$$E_{XC}^{hyb} = \alpha E_X^{HF} + (1-\alpha)E_X^{DFT} + E_C^{DFT} \qquad \text{(Eq. 2.52)}$$

Where E_{XC} is the exchange contribution and E_C is the correlation contribution; a is a coefficient generally fitted to large sets of experimental data. Two examples of hybrid functionals are B3LYP[49,50] and X3LYP.[51] These generally give improved results for aromatics when compared to pure DFT methods and because of this B3LYP is used within this work. A similar approach has previously been used to review the calculation of vibrational frequencies in MOFs by Keskin *et al*.[52]

2.4.4. Basis sets

The molecular oribtals of a system can be described as a set of functions, known as basis sets. Although increasing the size of these functions, the accuracy tends to increase as well; therefore a trade-off between accuracy and efficiency is required. One of the simplest basis sets, the Slater Type Orbital (STO-nG), describes each core or valence orbital using only one function; here n is an integer.

Basis sets based on Gaussian functions are also popular. This describes the molecular orbitals using a linear combination of atomic orbitals. These tend to have the expression:

$$\chi(x,y,z) = Nx^l y^m z^n e^{-\alpha r^2} \qquad \text{(Eq. 2.53)}$$

Here, x, y and z are Cartesian coordinates with l, m and n being the order of the function. a is the spread of the function and r is the distance from the nucleus. The sum of l, m and n determine the style of orbital present; for example when the sum of l, m and n equals zero, it corresponds to an s orbital – if it equals one it corresponds a p_x, p_y or p_z orbital. To improve the results the description of each orbital can be increased by either doubling or tripling the number of Gaussians used to describe the orbital – double-zeta and triple-zeta respectively. A split double-zeta basis set merely doubles the number of functions for the valance electrons. Including the Gaussians of the unoccupied orbitals can increase the

complexity of the basis set further and these are known as polarization functions.

Dunning *et al.* developed correlation consistent (cc) basis sets to try and reduce the computational expense by reducing the number of primitive functions required in a basis set.[53] It is possible to add diffusion functions to the Dunning basis sets so that weakly interacting molecules are more accurately described. An example of a correlation consistent basis set is cc-pVTZ; this is a correlation consistent, polarised Valence Triple Zeta basis set. When considering a periodic system, plane wave basis sets are often used; unlike other basis sets these are not atom centred.[2,53]

2.5. Molecular dynamics (MD)

2.5.1. General introduction

Molecular Dynamic (MD) simulations take an initial system and use this to generate a series of time-correlated points. This is achieved by using Newton's second law of motion to propagate a set of coordinates and velocities based on the systems initial configuration and velocities, based on a finite time steps; *force = mass x acceleration*. The time step used throughout this work is 0.0005 ps; this was deemed adequate when simulating these systems. The downside with this is that it means that a 10^6 simulation steps correlates to only 5 ns of real time. While this is far longer than is currently possible for DFT MD simulations, it can still exclude important phenomena. One advantage with *ab initio* MD is that no parameterisation is necessary. Also, it is possible to make and break bonds during the simulation.

In principle MD is deterministic, as if the same system was run with the same starting velocities, the trajectories should be identical. However, even minute differences in the initial coordinates of the atoms can generate vastly different trajectories. This is because the sheer number of time steps creates an exponential divergence between the two simulations. This can also be an issue if different computers are used, as these have different rounding errors – this would cause a disparity between two simulations. This means that although MD simulations are in principle deterministic in nature, in practice they can be considered non-deterministic, as they show chaotic behaviour after around 50 ps. However, the main reason these simulation are non-deterministic, is that the initial atomic velocities are randomly assigned. To run a MD simulation, the force on all the particles present within the system needs to be calculated; this is obtained by calculating the first derivative of the energy.

To calculate the trajectory of the particles within the system, the following equation must be solved:

$$\frac{d^2 r_i}{dt^2} = \frac{F}{m_i} \qquad \text{(Eq. 2.54)}$$

Where r_i and m_i are the position and mass of particle *i*, and *F* is the force. Rather than directly integrate the equations of motion, usually the finite differences method is used to determine the trajectory of the particles within the system. Here the integration is broken down into a series of small stages, all separated by a fixed interval of time; δt. This is achieved by first calculating the total force on the particles and from this the acceleration, *a*; the force at time *t* is calculated as a vector sum of the interactions of other particles. Using the current position of the particle, *r*, combined with the velocity, *v*, the acceleration can be used to determine new positions and velocities at $t + \delta t$. The force is assumed to remain as a constant and therefore it is possible to accurately describe how the system evolves with time; the accuracy of this description is dependant on the time step being sufficiently small. The forces acting on the particles in their new positions are then determined, which results in new positions are velocities at time $t + 2\delta t$ and this cycle is repeated until the simulation is completed. For these equations of motion, *r*, *v* and *a* are approximated as expansions of the Taylor series:

$$r(t+\partial t) = r(t) + \partial t v(t) + \frac{1}{2}\partial t^2 a(t) + \frac{1}{6}\partial t^3 b(t) + \frac{1}{24}\partial t^4 c(t) + \ldots \qquad \text{(Eq. 2.55)}$$

$$v(t+\partial t) = v(t) + \partial t a(t) + \frac{1}{2}\partial t^2 b(t) + \frac{1}{6}\partial t^3 c(t) + \ldots \qquad \text{(Eq. 2.56)}$$

$$a(t+\partial t) = a(t) + \partial t b(t) + \frac{1}{2}\partial t^2 c(t) + \ldots \qquad \text{(Eq. 2.57)}$$

$$b(t+\partial t) = b(t) + \partial t c(t) + \ldots \qquad \text{(Eq. 2.58)}$$

To initiate the calculations, the force of the atoms at $t=0$ must be known. This means velocities need to be assigned to the atoms and the most common approach is to assign these randomly, so that the Maxwell-Boltzmann distribution for that temperature is achieved.

The most common method for integrating the equations of motion is the Verlet Algorithm.[34] This uses the positions and accelerations of the atoms at one time

step, *t*, in combination with the position of the previous time step, *r(t-δt)* to calculate the new position of the particle at *t+δt*; *r(t+δt)*. This can be written as:

$$r(t+\partial t) = r(t) + \partial t v(t) + \frac{1}{2}\partial t^2 a(t) + \ldots \quad \text{(Eq. 2.59)}$$

$$r(t-\partial t) = r(t) - \partial t v(t) + \frac{1}{2}\partial t^2 a(t) + \ldots \quad \text{(Eq. 2.60)}$$

Subsequently, these equations can be added together to give the following expression:

$$r(t+\partial t) = 2r(t) - r(t-\partial t) + \partial t^2 a(t) \quad \text{(Eq. 2.61)}$$

As the velocities are not contained within this equation, these need to be calculated by a different manner. Two common ways are used; one where the velocities can simply be calculated by determining the difference in the positions at time *t+δt* and *t-δt*:

$$v(t) = \left[r(t+\partial t) - r(t-\partial t)\right]/2\partial t \quad \text{(Eq. 2.62)}$$

Another approach is to calculate the velocities for the atoms at the half-step, *t+½δt*:

$$v\left(t+\frac{1}{2}\partial t\right) = \left[r(t+\partial t) - r(t)\right]/\partial t \quad \text{(Eq. 2.63)}$$

To implement the Verlet Algorithm, only two sets of positions are required, along with the acceleration, higher derivative terms are not required and this means storage requirements are reasonable. One major disadvantage is that the velocities themselves are never explicitly calculated therefore this can lack precision. In combination with this, the algorithm is not self-starting and random initial velocities have to be assigned to the atoms; the velocity can only be calculated when the positions of the following step have been calculated.

There are several variations of the Verlet algorithm, including the leapfrog algorithm and the velocity Verlet algorithm.[34] These aim to explicitly describe the velocities during the simulations. Considering the leapfrog algorithm first, it uses the following expressions:

$$r(t+\partial t) = r(t) + \partial t v\left(t+\frac{1}{2}\partial t\right) \quad \text{(Eq. 2.64)}$$

$$v\left(t+\frac{1}{2}dt\right)=v\left(t-\frac{1}{2}\partial t\right)+\partial t a(t) \quad \text{(Eq. 2.65)}$$

When using the leapfrog algorithm, the velocities at t+½δt are calculated using the velocities at *t-½δt* in combination with the acceleration. Equation 2.65 allows the positions *r(t+δt)* to be calculated using the velocities just calculated together with the positions at time t. It is also possible to calculate the velocities at time *t*:

$$v(t)=\frac{1}{2}\left[v\left(t+\frac{1}{2}\partial t\right)+v\left(t-\frac{1}{2}\partial t\right)\right] \quad \text{(Eq. 2.66)}$$

The name of the algorithm is quite apt as the velocities 'leapfrog' over the positions to give new values at *t+½δt* to give their values. This process is then repeated for the positions such that they jump over the velocities to give new values at *t+δt*; this process is constantly cycled. The two main advantages of using the leapfrog algorithm over the Verlet algorithm are that the velocities are explicitly used, and the difference of large numbers does not need to be calculated. However, the main disadvantage is that the positions and velocities are not in synchronization; this prohibits the kinetic energy contribution to the total energy from being calculated. This is required to determine the potential energy of the system.

The velocity Verlet tries to combat this by giving the positions, velocities and accelerations all at the same time, while not compromising on precision. This is shown in the following expressions:

$$r(t+\partial t)=r(t)+\partial t v(t)+\frac{1}{2}\partial t^2 a(t) \quad \text{(Eq. 2.67)}$$

$$v(t+\partial t)=v(t)+\frac{1}{2}\partial t\left[a(t)+a(t+\partial t)\right] \quad \text{(Eq. 2.68)}$$

$$v\left(t+\frac{1}{2}\partial t\right)=v(t)+\frac{1}{2}\partial t a(t) \quad \text{(Eq. 2.69)}$$

$$v(t+\partial t)=v\left(t+\frac{1}{2}\partial t\right)+\frac{1}{2}\partial t a(t+\partial t) \quad \text{(Eq. 2.70)}$$

This needs to be implemented as a three-stage procedure, as the accelerations at both *t* and *t+δt* are required to deduce the new velocities. These steps are:

1. The positions at time *t+δt* are calculated using Equation 2.67, using the velocities and accelerations at time *t*. The velocities can then be

determined at time $t+½\delta t$ using Equation 2.69;

2. Next, new forces can be computed for the current positions, giving $a(t+\delta t)$;
3. Finally, the velocities at time $t+\delta t$ are determined using Equation 2.70.

It is important to choose the correct integration algorithm for the simulation in question; this should be fast, require minimal memory and be easy to program. That said, different algorithms perform better as the length of the time step is increased. Therefore although one algorithm may be more computationally more expensive in a like-for-like MD simulation when compared to another algorithm, if the time step can be increased without loss of accuracy, it is in fact more efficient. Therefore consideration of the time step used is important. Throughout this work, the leapfrog integration algorithm has been used.

2.5.2. Time step

The time step is often the rate-limiting factor when it comes to the speed of a MD simulation. It is important to strike a balance between efficiency and accuracy. If the time step is too small, only a limited portion of the phase space will be sampled, whereas too large a time step would lead to instabilities in the integration algorithm – this is mainly caused by the overlap of atoms; Figure 2.17. If the time step is too large, atoms could be placed on top of one another and this would lead to a large increase in energy. Both these instabilities would lead to the simulation crashing, as there would a violation of energy and linear momentum conservation.

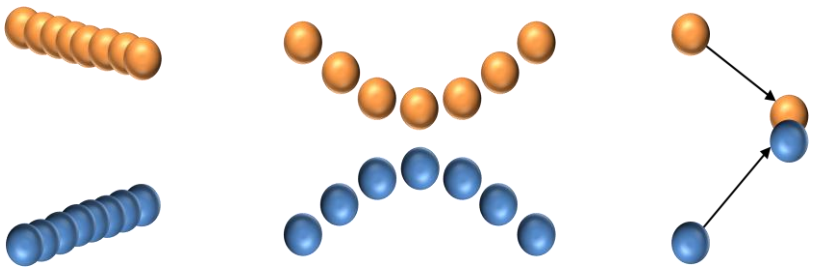


Figure 2.17. It is important to choose the correct time step, so that the phase space is covered both efficiently and smoothly (centre). If too small a time step is chosen (left) the phase space is covered too slowly. If the time step is too large, the phase space is covered to quickly and therefore atoms can crash into one other, which is energetically unstable. Figure adapted from reference 2.

2.5.3. Molecular dynamic ensembles

As mentioned earlier, MD simulations rely on the integration of Newton's equations of motion and therefore sample a microcanonical ensemble by default; this is constant energy. It is important to ensure compatibility with experimental results, and therefore it is desirable to instead sample configurations from a canonical ensemble – one where the temperature is kept constant. To do this a thermostat is required.[54] There are a number of different ensembles used for both MC and MD simulations. MD simulations can use a NVE ensemble; where the number of particles (N), the volume, (V), and the total energy, (E), remain constant. The NVE microcanonical ensemble can be obtained by varying the temperature and pressure, so that Newton's equation is satisfied. This is used for isolated systems, and means that the every microstate has equal probability. It is possible to change the fixed conditions in an ensemble, so that different properties of the systems can be examined. For example the NVT ensemble keeps the number of particles, volume and temperature constant and by varying the pressure and total energy of the system Newton's equation can be solved. Another alternative is the NPT ensemble; this is similar to the NVT ensemble, but instead of the volume being kept constant, the pressure is fixed and the volume in combination with the total energy of the system now varies. For both the NVT and NPT ensembles, the temperature is kept at a constant. This is achieved by coupling the system to a heat bath; this is fixed at the correct temperature so that it can add or remove heat to the system to maintain the temperature constant - this is achieved by using a thermostat.[55] This method can be extended to controlling the pressure of the system by using a barostat; this creates an external pressure bath that can exchange volume. Both the Nose-Hoover thermostat and barostat are used throughout this work.[55]

2.6. Monte Carlo simulations (MC)

Monte Carlo (MC) techniques are not deterministic techniques meaning that it is potentially possible for an entire region of phase space to be sampled. A sequence of points in phase space is generated from an initial geometry by adding an external force to the coordinates of an atom chosen randomly within the system.

The specific MC method used throughout this work is the Metropolis algorithm.[2] This is used to generate a sequence of configurations using the initial geometry of the system and subsequently a random force is given to the coordinates of a given particle (an atom or molecule) which itself is chosen arbitrarily. This new

configuration is accepted if the overall energy of the system is lower than that of the initial configuration. The main advantage of MC is that totally unconnected configurations are generated, making it possible to in theory sample large regions of phase space. The Metropolis method also accepts some configurations even if the energy is greater than the original; this is aimed to simulate going uphill on the PES. These configurations though are only accepted with a probability of the Boltzmann factor, $\beta = e^{\frac{-\Delta v}{K_B T}}$, meaning that there is a lower probability of these configurations being sampled. Here, v is the potential energy of the system, k_B is the Boltzmann constant and T is the temperature. This allows the ability of escaping local minima. This method also ensures that the variables within the ensemble obey the Boltzmann distribution. Including this inclusion criterion means that to have a reasonable acceptance ratio the step size has to be relatively small and therefore simulations using several million MC steps are required; generally a desirable ratio of 0.5 is used. A MC step can be a translation, rotation (unless the molecule is spherical), insertion or deletion. Consequently, even with this many steps, the deviation from the initial starting geometry is minimal.

One particular advantage about using the Metropolis algorithm is that it is only dependent on the configuration immediately previous, thus generating a Markov chain of states; this makes sure that the outcome of each iteration is only dependent on the preceding iteration. In addition to this, each iteration only has a finite set of possible outcomes. A flow chart outlining the procedure involved in the Metropolis Monte Carlo algorithm can be seen in Figure 2.18.

To simulate the system at constant chemical potential, μ, temperature, T, and volume, V, Grand Canonical Monte Carlo (GCMC) can be used.[2] It is also possible to perform a GCMC calculation at constant activity, z. This is related to μ using the following equation:

$$\mu = k_B T \ln \Lambda^3 z \qquad \text{(Eq. 2.71)}$$

Where Λ is the de Broglie wavelength, given by:

$$\Lambda = \sqrt{h^2 / 2\pi m k_B T} \qquad \text{(Eq. 2.72)}$$

It is also possible to calculate the pressure directly from μ, using the ideal gas law:

$$P = \left[\exp\left(\mu / k_B T\right) k_B T\right] / \Lambda^3 \qquad \text{(Eq. 2.73)}$$

One major advantage of using GCMC is that it allows the number of particles within the system to vary throughout the simulation; this is performed using the following steps:

1. Displacement of particles using the Metropolis method
2. Destruction of a particle
3. A particle is created at a random position within the simulation box.

to determine whether to accept or reject the destruction or creation of a particle. To calculate this, the following is calculated:

$$\Delta D = \frac{\left[\upsilon_{new}\left(r^N\right) - \upsilon_{old}\left(r^N\right)\right]}{k_B T} - \ln\left(\frac{N}{zV}\right) \qquad \text{(Eq. 2.74)}$$

Where $v_{new}(r^N)$ is the potential energy function. To determine whether the creation of a particle is accepted, the following is calculated:

$$\Delta C = \frac{\left[\upsilon_{new}\left(r^N\right) - \upsilon_{old}\left(r^N\right)\right]}{k_B T} - \ln\left(\frac{N}{zV+1}\right) \qquad \text{(Eq. 2.75)}$$

The moves are accepted if ΔD or ΔC is negative. If however ΔD or ΔC is positive, the exponential of the either $\exp(-\Delta D / k_B T)$ or $\exp(-\Delta C / k_B T)$ is calculated, and subsequently compared to a random number between 0 and 1, as detailed earlier.

It is also important that the probability to create and destroy of a particle is equal. The ratio of create/destroy moves to translate/rotate moves can vary, however convergence is generally achieved more rapidly when all these have approximately equal frequency. Consequently, the ratio of the four options was kept equal during this work.

Recently, GCMC simulations have been used for simulating both adsorption isotherms[56,57] and the sequential loading of molecules.[58] Within this work, the Sorption and Adsorption locator modules, found with the Materials Studio 5.0 software package were used.

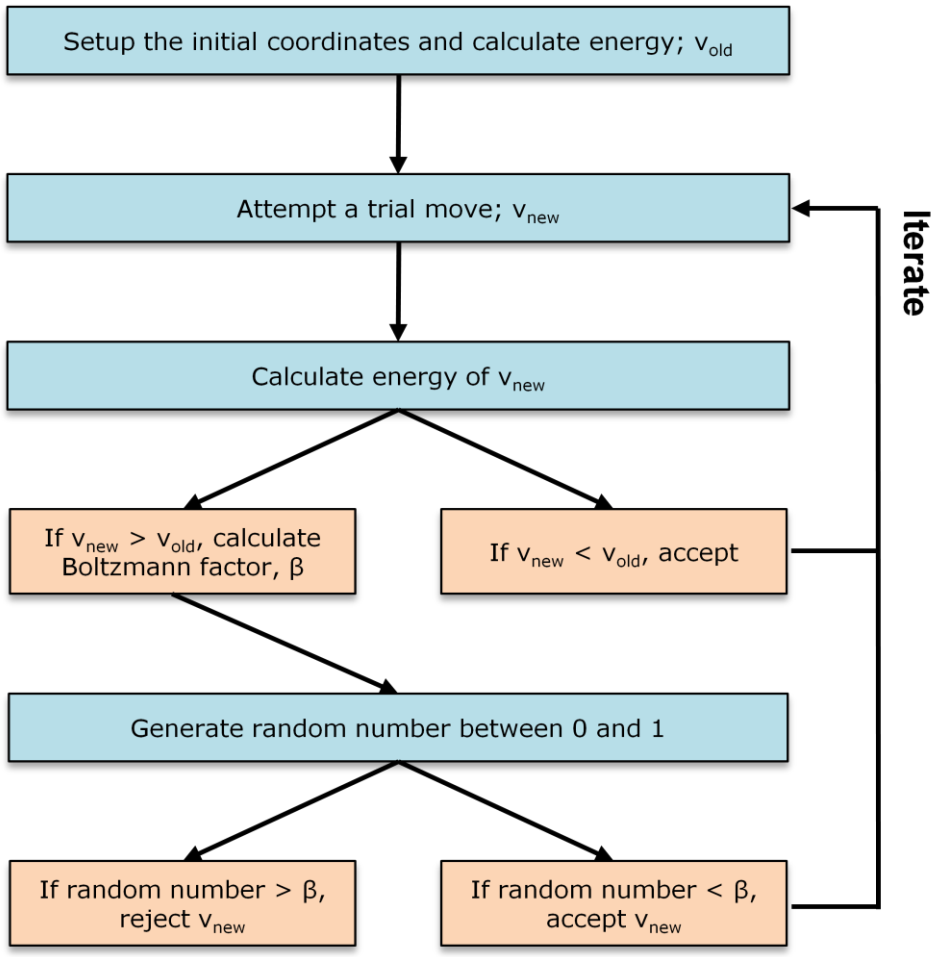


Figure 2.18. Flow through example of the Metropolis algorithm performing a MC simulation.

2.7. Computational software used

2.7.1. Gaussian09 package

DFT calculations were used to generate PES landscapes for all the intramolecular degrees of freedom within the cage fragment. These were obtained using B3LYP calculations with a 6-31G(d,p) basis set within the Gaussian09 package.[59]

2.7.2. DL_POLY_2.20 MD package

All MD simulations simulations were carried out using DL_POLY 2.20; Smith, Forester and Todorov developed this at Daresbury Laboratory.[60] DL_POLY_prep, developed by David Willock,[61] was used to convert the starting configurations into the correct file formats. Analyse_hist[61] was used to examine the MD trajectories; this calculated the relative energies, deviations of the intramolecular degrees of freedom, mean squared displacement of gas molecules and from this their diffusion coefficients.

2.7.3. Accelrys materials studio 5.0

Materials Studio was used to visualise the structures as well as calculate both the surface areas. Both the Forcite and Discover modules were used to minimise any structures; here CSFF was implemented. The Atom Volumes and Surfaces module was used as one method to generate the calculated surfaces.[18]

2.7.4. Visual molecular dynamics

The Visual Molecular Dynamics, VMD, was used to display and analyse the trajectories of all MD calculations.[62] It was also used in combination with Zeo^{++} to generate both the connectivity pore topologies as well as accessible surface area plots.[21]

2.7.5. Zeo++

Zeo^{++}, developed by Haranczyk, was used in combination with VMD to visualise both the connectivity and pore topologies of the systems. In addition to this, Zeo^{++} was used to generate binary cube files that were in turn read into the VISIT software package,[63] so that any dynamic connectivity could be studied.

2.8. References

(1) West, A. R. *Basic Solid State Chemistry*, Wiley, **1988**.

(2) Leach, A. R. *Molecular Modelling: Principles and Applications,* Prentice-Hall, **2001**.

(3) Kolasinski, K. K. *Surface Science: Foundations of Catalysis and Nanoscience*, Wiley, **2012**.

(4) Panella, B.; Hirscher, M. Adv. Mater. **2005**, *17*, 538–541.

(5) Fletcher, A. J.; Thomas, K. M.; Rosseinsky, M. J. J. of Solid State Chem. **2005**, *178*, 2491–2510.

(6) Nouar, F.; Eckert, J.; Eubank, J. F.; Forster, P.; Eddaoudi, M. J. Am. Chem. Soc. **2009**, *131*, 2864–2870.

(7) Roque-Malherbe, R. M. A. *Adsorption and Diffusion in Nanoporous Materials*; CRC Press, **2010**.

(8) Donohue, M. D.; Aranovich, G. L. Adv. Colloid Interface Sci. **1998**, *76*, 137–152.

(9) Langmuir, I. J. Am. Chem. Soc. **1916**, *38*, 2221–2295.

(10) Skoulidas, A. I. J. Am. Chem. Soc. **2004**, *126*, 1356-1357.

(11) Skoulidas, A. I.; Sholl, D. S. J. Phys. Chem. B. **2002**, *106*, 5058–5067.

(12) Skoulidas, A. I.; Sholl, D. S. J. Phys. Chem. B. **2005**, *109*, 15760–15768.

(13) Skoulidas, A. I.; Sholl, D. S. J. Phys. Chem. A. **2003**, *107*, 10132–10141.
(14) Connolly, M. L. Science. **1983**, *221*, 709–713.
(15) Pignataro, B. *Molecules at Work*, Wiley, **2012**.
(16) http://www.ccdc.cam.ac.uk/products/csd/radii.
(17) Jensen, F. *Introduction to Computational Chemistry*, Wiley, **2013**.
(18) Accelrys; Accelrys Software, Inc.: San Diego, **2011**.
(19) Shrake, A.; Rupley, J. A. J. Mol. Biol. **1973**, *79*, 351–371.
(20) Do, D. D.; Herrera, L. F.; Do, H. D. J. Colloid Interface Sci. **2008**, *328*, 110–119.
(21) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Microporous Mesoporous Mater. **2012**, *149*, 134–141.
(22) Duren, T.; Millange, F.; Férey, G.; Walton, K. S.; Snurr, R. Q. J. Phys. Chem. C. **2007**, *111*, 15350–15356.
(23) Herrera, L. F.; Do, D. D.; Birkett, G. R. Journal of Colloid and Interface Science. **2008**, *320*, 415–422.
(24) Walton, K. S.; Snurr, R. Q. J. Am. Chem. Soc. **2007**, *129*, 8552–8556.
(25) Rowsell, J. L. C.; Yaghi, O. M. J. Am. Chem. Soc. **2006**, *128*, 1304–1315.
(26) Robeson, L. M. J. Mem. Sci. **1991**, *62*, 165–185.
(27) Thomson, K. T.; Gubbins, K. E. Langmuir. **2000**, *16*, 5761–5773.
(28) Düren, T.; Bae, Y.-S.; Snurr, R. Q. Chem. Soc. Rev. **2009**, *38*, 1237–1247.
(29) Sarkisov, L.; Harrison, A. Mol. Sim. **2011**, *37*, 1248–1257.
(30) Martin, R. L.; Smit, B.; Haranczyk, M. J. Chem. Inf. Model. **2012**, *52*, 308–318.
(31) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Chem. Rev. **1994**, *94*, 1887–1930.
(32) Warshel, A. J. Chem. Phys. **1970**, *53*, 582–594.
(33) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. J. Am. Chem. Soc. **1989**, *111*, 8551-8566.
(34) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, Oxford Science Publications, **1989**.
(35) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. J. Am. Chem. Soc. **1996**, *118*, 11225–11236.
(36) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. J. Comput. Chem. **1986**, *7*, 230–252.
(37) Nilsson, L.; Karplus, M. J. Comput. Chem. **1986**, *7*, 591–616.
(38) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. J. Am. Chem. Soc. **1979**, *101*, 6842–6852.

(39) Allinger, N. L. J. Am. Chem. Soc. **1977**, *99*, 8127-8134.
(40) Meier, R. J.; Maple, J. R.; Hwang, M. J. J. Phys. Chem. **1995**, *99*, 5445-5456.
(41) Sun, H.; Mumby, S. J.; Maple, J. R. J. Am. Chem. Soc. **1994**, *116*, 2918-2981.
(42) Sun, H. J. Comput. Chem. **1994**, *15*, 752–768.
(43) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. J. Comput. Chem. **2004**, *15*, 162–182.
(44) Kohn, W.; Sham, L. J. Phys. Rev. **1965**, *140*, A1133.
(45) Becke, A. Phys. Rev. A. **1988**, *38*, 3098–3100.
(46) Lee, C.; Yang, W.; Parr, R. Phys. Rev. B. **1988**, *37*, 785–789.
(47) Henderson, T. M.; Janesko, B. G.; Scuseria, G. E. J. Chem. Phys. **2008**, *128*, 194105.
(48) Becke, A. D. J. Chem. Phys. **1993**, *98*, 5648–5652.
(49) Kim, K.; Jordan, K. D. J. Phys. Chem. **1994**, *98*, 10089–10094.
(50) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. J. Phys. Chem. **1994**, *98*, 11623–11627.
(51) Xu, X.; Goddard, W. A. Proc. Natl. Acad. Sci. U.S.A. **2004**, *101*, 2673–2677.
(52) Keskin, S.; Liu, J.; Rankin, R. B.; Johnson, J. K.; Sholl, D. S. Ind. Eng. Chem. Res. **2009**, *48*, 2355–2371.
(53) Dunning, T. H. J. Chem. Phys. **1989**, *90*, 1007.
(54) Hünenberger, P. H. Adv. Polym. Sci. **2005**, *173*, 105–149.
(55) Melchionna, S.; Ciccotti, G.; Lee Holian, B. Mol. Phys. **1993**, *78*, 533–544.
(56) Del Regno, A.; Siperstein, F. R. Microporous Mesoporous Mater. **2013**, *176*, 55–63.
(57) Thornton, A. W.; Furman, S. A.; Nairn, K. M.; Hill, A. J.; Hill, J. M.; Hill, M. R. Microporous Mesoporous Mater. **2013**, *167*, 188–197.
(58) Yang, Z.; Peng, X.; Cao, D. J. Phys. Chem. C. **2013**, *117*, 8353–8364.
(59) Gaussian 09 Software Program.
(60) Smith, W.; Forester, T. R. J Mol Graph. **1996**, *14*, 136–141.
(61) http://theory.chem.cf.ac.uk/~dave/soft_intro.html.
(62) http://www.ks.uiuc.edu/.
(63) https://wci.llnl.gov/codes/visit/home.html.

Chapter 3

Bespoke force field for simulating the molecular dynamics of porous organic cages

3.1 Introduction

It has been shown that it was possible to synthesise tetrahedral organic cages that form porous molecular crystals which adsorb small guest molecules, such as gases.[1,2] The porosity of these materials was a result of the inefficient crystal packing of these discrete cage subunits. This was aided by the intrinsic, covalently prefabricated, molecular voids found within the cages. The pore structure and connectivity can be directed by both the functional groups attached to the cage vertices and by the solvent from which they were recrystallised, such that it was possible to connect or disconnect the cage cavities within an extended crystal structure. This provides a new design principle for preparing highly porous organic materials by the synthetically-directed assembly of prefabricated molecular pores.[3] In contrast to metal-organic polyhedra,[4-6] these cage molecules are wholly organic and composed entirely of covalent C-C, C-H, N-H and C-N bonds. They are also soluble in common organic solvents.

In this study, we show that the structure and flexibility of these organic imine cages was not well represented by "off the shelf" force fields (FF). Hence, we develop and test a bespoke FF for simulating the molecular dynamics of a series of porous organic imine cage materials. Our aim was to develop a single FF encompassing the intramolecular FF terms required to describe the cage structure, along with the intermolecular interactions between the molecules that determine the crystal packing. It was important that the FF was transferable over a number of cage systems and hence the FF was fitted using single X-ray crystal data for four known structures: two polymorphs of **CC1** (**CC1α** and **CC1β**)[7] **CC2**,[1] and **CC3**.[1] Furthermore, the FF was then tested for the crystal structure of a recently-synthesised member of this family of cage molecules, **CC4**,[8] formed from the reaction of (*R*,*R*)-1,2-diaminocyclopentane with 1,3,5-triformylbenzene. In addition, the cocrystal of two cages (**CC1α**, **CC3**),[3] was used as a secondary check on the performance. This new FF will henceforth be referred to as the 'cage specific force field' (CSFF).

3.2 Simulation Details

3.2.1. Minimisation

The Smart minimiser within Materials Studio 5.0 was used for minimisation of the cages in both the gas phase and their crystal structures. This combines the steepest descent, conjugate gradient and Newton-Raphson methods in a cascade. The steepest descent method was used first with a convergence criterion of 1000

kcal mol^{-1} $Å^{-1}$. The conjugate gradient method was then used with the Fletcher-Reeves algorithm, with a convergence criterion of 10 kcal mol^{-1} $Å^{-1}$. Finally, the BFGS algorithm was used for the Newton-Raphson method. A much more accurate convergence of $1x10^{-5}$ kcal mol^{-1} $Å^{-1}$ was used. The maximum amount of atoms allowed in each minimisation was 200; if this was exceeded, the Smart minimiser removed this method from the cascade, as it becomes computationally too expensive, and the convergence criterion of the conjugate gradient method was set to $1x10^{-5}$ kcal mol^{-1} $Å^{-1}$.

3.2.2. Similarity comparisons

To distinguish the similarity between both gas phase and crystalline systems, the Field method available in Materials Studio 5.0 was used. This aligns the structures by calculating both the steric and electrostatic fields around each system, and subsequently re-orientates one system so that the fields overlap. The steric and electrostatics contributions can be weighted and for this work this was weighted such that only the steric interactions were considered. The steric field was defined as the potential for each system seen by a probe carbon atom. The similarity of the structures was then calculated, with a similarity of 1 indicating a perfect match.

3.2.3. Molecular dynamics setup

Gas phase MD simulations were carried out for **CC1**, **CC2** and **CC3**, using the Discover module from Materials Studio, for 1 ns with an NVT ensemble at 298 K sampled every 0.5 ps. The X-ray single crystal structure was used as a starting point, and for simplicity only one positional isomer of **CC2** was used.[1] The *2220* isomer was chosen, as this was the isomer most likely to be found in the crystal structure based on the partial occupancies. In reality, the crystal structure for **CC2** contains greater disorder due to positional isomers based on placement of the vertex methyl groups. This is explained further in Appendix A1.

DL_POLY 2.20 was used to run MD simulations to investigate the on the dynamic behaviour of the cage structures. For this, supercells of the four cage systems, **CC1α**, **CC1β**, **CC2**, and **CC3**, were generated from their respective crystal structures, and an MD simulation for each system was ran using an NPT ensemble at 298 K and a pressure of 1 atm. A timestep of 0.5 fs was used, and the systems initially underwent a 50 ps equilibration, followed by a subsequent production run for 1 ns. The reported data was averaged across the production period of the simulations only. The cross-terms (bond-bond, bond-angle, etc.)

that are used in the full equation set for PCFF potential types have been omitted from the MD simulations, as the required functional forms are not implemented in DL_POLY 2.20. This means that comparison of the MD simulations, to the energy minimisations, carry the assumption that the absence of these cross-terms has a negligible effect on the results.

3.3. Testing existing FFs

To see whether any off-the-shelf FFs were capable of describing our cages, three well-known FFs were used to minimise the cage structures, and several internal degrees of freedom were evaluated. Initial tests showed that generalised FFs, such as the Universal force field[9] (UFF), the polymer consistent force field (PCFF),[10] and COMPASS,[11,12] do not accurately simulate the structures; Table 3.1.

FF Used	C-N-C (°)	N-C-C (°)	N-C-C-N (°)	Similarity (%)	RMS atom distance (Å)
Crystal Structure	115.50	108.84	69.48	Target	Target
Universal	123.31	110.00	62.67	98.80	0.073
Compass	122.01	116.67	112.74	93.01	0.237
PCFF	117.75	117.07	62.67	98.73	0.078

Table 3.1 – **CC1** was minimised in the gas phase using a variety of generic FFs and compared to the X-ray crystal structure data. For similarity and RMS atom distance, hydrogens have been omitted.

As an example, a comparison of the optimised structures of a cage of **CC1** in the gas phase from each FF with the initial crystal structure geometry is given in Table 3.1. The Universal FF overestimates the C-N-C angle for the imine group by around 8°, and shows a root mean squared (RMS) distance between the optimised and reference structure of 0.073 Å. The COMPASS FF shows a similar overestimation of the C-N-C angle for the imine group and the N-C-C-N torsion was over 40° greater than that experimentally observed, resulting in an RMS distance between calculated and observed structures of 0.237 Å. PCFF does give the C-N-C angle for the imine group to within 3°, but the N-C-C angle associated with the imine group was significantly over estimated. Even so, the RMS distance of atoms in the PCFF[10] structure from the experimental structure was only 0.078 Å. In particular, the planarity of the region around the phenyl rings of the cages, which depends on conjugation with the imine groups,[1] was poorly represented, Figure 3.1.

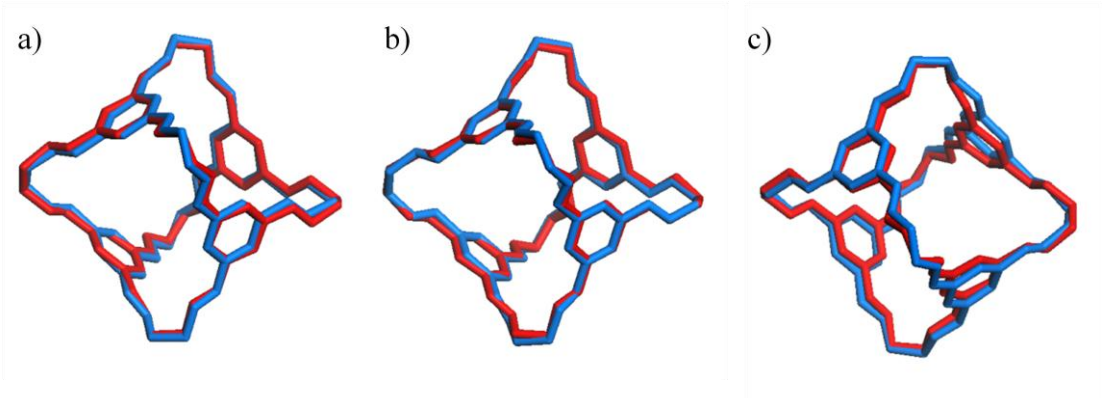


Figure 3.1 – Figure showing the similarity to crystal structure (red) post FF minimisation (blue) using a) UFF, b) PCFF and, c) COMPASS FF.

PCFF[10] lacked an accurate parameterisation of the imine group in the environment found in these cage materials. Nonetheless, PCFF did describe the porous cage molecules more accurately than the other off-the-shelf FFs we tested, and so we decided to use this as the basis for a new, bespoke FF that would accurately describe a range of discrete porous imine cage molecules.

		Cell Parameters							
System	Supercell size	a / Å	b / Å	c / Å	α / °	β / °	Υ / °	Density (g/cm^3)	No. atoms
CC1α	4 x 4 x 2	51.24	43.64	73.62	90.00	97.49	90.00	1.03	13824
CC1β	4 x 2 x 4	84.92	42.46	42.48	90.00	90.00	120.00	0.93	10368
CC2	4 x 4 x 4	75.08	75.08	43.68	90.00	90.00	120.00	0.87	16128
CC3	2 x 2 x 2	49.60	49.60	49.60	90.00	90.00	90.00	0.97	10752
CC4	2 x 2 x 4	47.85	47.85	43.83	90.00	90.00	120.00	0.95	7200
CC1α, CC3	2 x 2 x 2	48.55	48.55	48.55	90.00	90.00	90.00	0.89	8832
Benzene	3 x 3 x 3	21.86	27.60	20.06	90.00	90.00	90.00	1.16	1296
Aniline	2 x 8 x 4	43.64	46.94	33.55	90.00	101.01	90.00	1.17	7168

Table 3.2 – Table listing the lattice parameters and number of molecules used for all MD simulations.

The main issue seems to be the imine functionality, which is directly attached to an aromatic ring. For example, Figure 3.2 shows a whole subset of different molecules parameterised within the COMPASS FF. Here, it was evident that benzene has been accurately parameterised, as well as alkanes, amines, and amides; imines, though, were omitted. This would explain why the aromatic region of the system seems to be well produced, while the imine regions were not. UFF was fitted for a wide range of functional groups; unstrained and uncongested hydrocarbons, saturated ethers and phosphines, alkenes, silanes, saturated amines, aromatic systems, and simple unconjugated multiple bond containing compounds such as nitriles, ketones, and imines.[9] However, like the COMPASS FF, the planarity of the phenyl ring in combination with an imine bond

was not present in the parameterisation criteria. Although this was also true of PCFF, it was a useful starting point as it has been used to accurately simulate polymers and organic materials and hence we can develop the additional PCFF parameters through FF fitting to density functional theory (DFT) reference data for the absent functionalities.

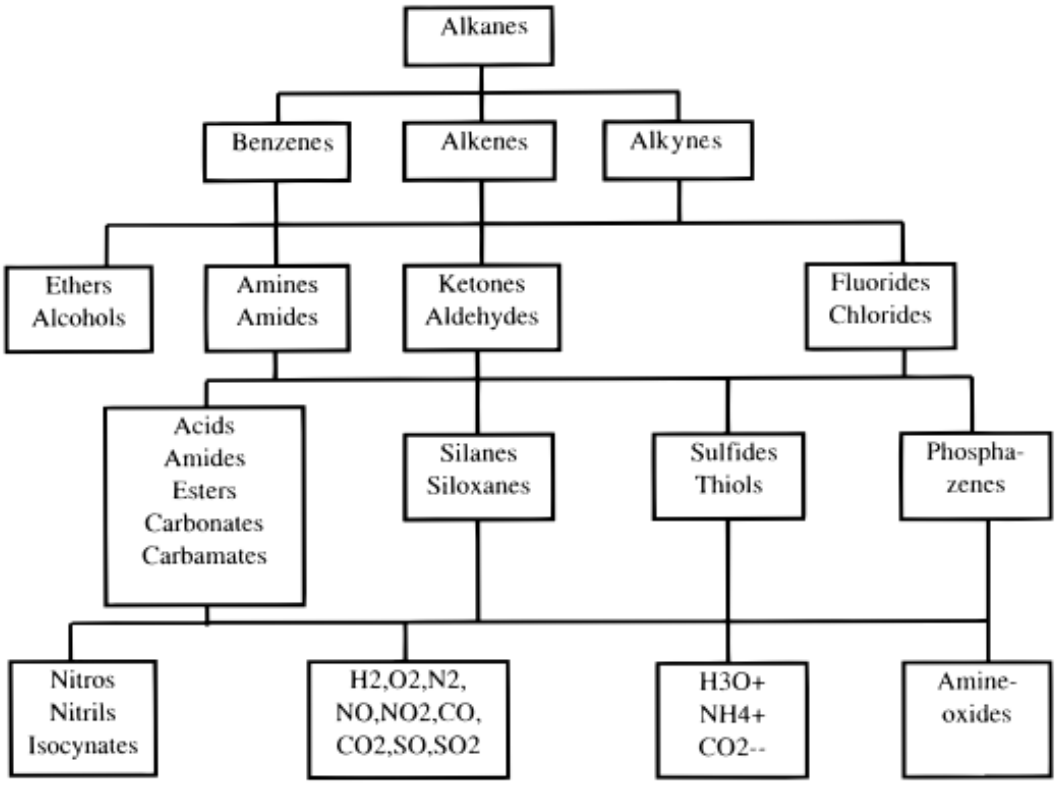


Figure 3.2 – Parameterisation precedence tree for COMPASS FF. The parameterisation starts from the top and all parameters determined at one level were fixed and transferred as many times as possible to the next level. Figure taken from reference 12.

The aim of this FF was to accurately describe **CC1**, **CC2**, and **CC3**. The only difference between these cges was the functionality of the vertices, meaning an unfunctionalised 108-atom cage core (**CC1**) acts as a good description of all three cages. Therefore, *ab initio* data was fitted to describe this cage core and the accuracy for the other cages can be subsequently analysed to see whether this assumption was appropriate.

When using MD, it was important to use a system that was large enough so that periodic boundaries do not affect the results. Therefore, supercells of crystal structures were generally used. The **CC1α** unit cell has dimensions *a* = 12.81 Å, *b* = 10.91 Å, *c* = 24.54 Å, α= 90.00°, β= 97.49° and γ= 90.00°. This was used to generate a 4 × 4 × 2 supercell which contains 128 cages with a total of 13,824 atoms. As this system consists of 128 identical cages, one was extracted and

studied to see if there were any degrees of symmetry within the molecule. It was clear that a 24 atom fragment was repeated around the cage system four times and this describes all the intramolecular parameters required to describe a whole cage. This meant that only 34 intramolecular potentials could describe the whole cage system. To validate this assumption, after the fragment was fully parameterised, a whole cage was minimised and the accuracy of the FF tested by comparing to single crystal structure data. This procedure is shown is Figure 3.3.

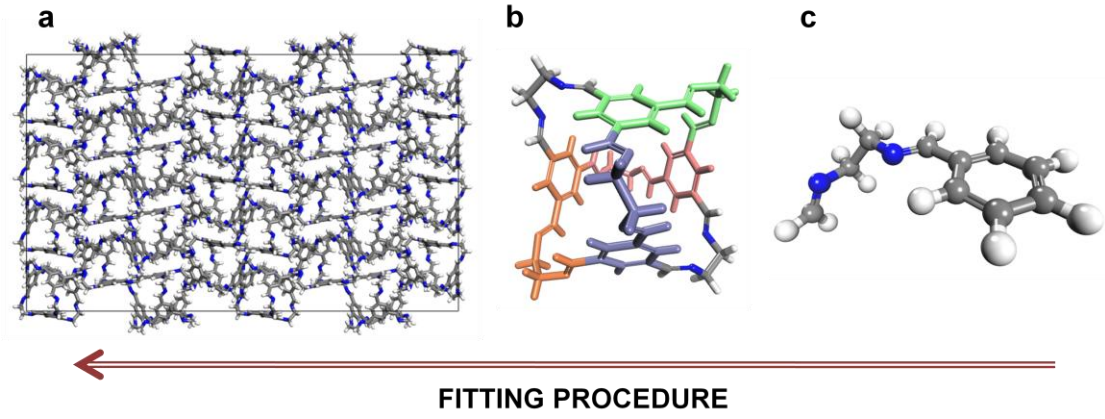


Figure 3.3 – A 4 x 4 x 2 supercell of the crystal structure of **CC1α**; containing 13.824 atoms. This contains a 108 atom repeat unit; this was a single cage. A common fragment of 24 atoms was identified within a single cage and this was extracted from the crystal structure. This fragment was then used to fit the parameters required for the intramolecular degrees of freedom. The 24-atom fragment highlighted four times in the single cage to show how it was constantly repeated across the cage core.

3.4. Identification of absent parameters

The adapted ZEBEDDE[13] code* read in the structure and subsequently identified the atomic connectivity, based on atom-atom cutoffs. This can then be used to identify all the bond stretch, angle bend, and dihedral sets within the molecule. Consequently, this then matched the potentials available from the FF file, and if any were unaccounted for, their contribution to the energy as zero, thus highlighting them as being absent. Inspection of the FF parameter file using this adapted version of the ZEBEDDE code showed that the PCFF did not contain parameter types for the imine group in the environment found within these cage structures. This meant that equivalences from within the FF were assumed, though this assumed transferability appears to have led to the inaccuracies in the simulated structures. An equivalency is when an existing FF parameter was used to represent a parameter not found within the FF. One such example of this was

* This was adapted by Dr David Willock to identify the missing parameters within a FF.

the imine bond length; this parameter was absent within PCFF parameter file and thus defaulted to using the parameter to describe an amine bond – this creates an overly long equilibrium bond of 1.47 Å opposed to the more reasonable 1.27-1.30 Å.

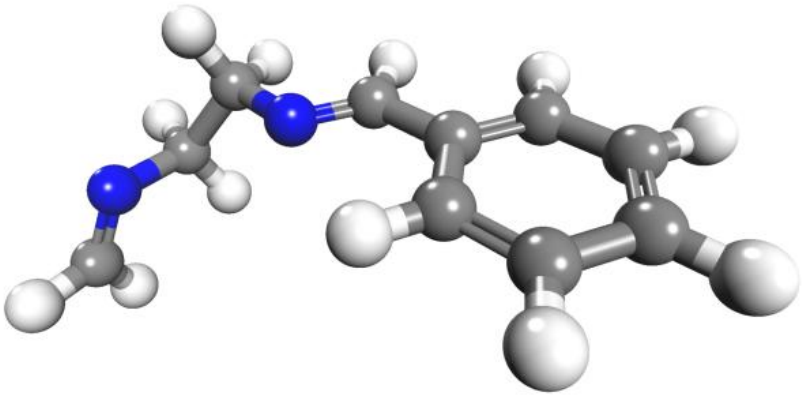


Figure 3.4 – Picture of fragment used to parameterise CSFF.

To prevent incorrect equivalences being used, each atom within the system was given a FF atom type, so that potential parameters could be described explicitly. The molecular fragment chosen to generate PESs for all the parameter sets is displayed in Figure 3.4. Due to the symmetry of the cage molecules, this fragment is repeated four times for each cage. All the aromatic carbons located on the cage face were denoted *"cp"*, the carbon atoms found within the amine double bond were denoted *"c=1"*, while the aliphatic carbons have the FF atom type *"c"*. The hydrogens were all given the atom type *"h"* while the amine nitrogen has an *"n="* FF atom type. These atom types can be further extended, as although they described the same intramolecular and intermolecular parameters, their local environment dictates that their partial charges were different. For example, different partial charges on the *cp* atom types generate different FF atom types. These were given the assignments *"cp_1"* and *"cp_2"*; *cp_1 is* connected to the aliphatic cage vertices, whereas *cp_2* was attached to one of the aromatic hydrogens. Additionally, the hydrogens also have atom type subsets; *hc_1* or *hc_2*; *hc_1* describe the hydrogens situated on the aliphatic cage vertices, whereas *hc_2* describe the aromatic hydrogens. As the 24-atom fragment contains only a small subset of all the atoms in the system, only 32 parameters were required to describe the intramolecular forces of the system; 8 bonds, 10 angles and 14 dihedrals. The FF atom types are listed in Table 3.3.

FF atom type	Description
c	Generic sp^3 hybridised carbon
cp	sp^2 hybridised aromatic carbon
c=1	Non aromatic end doubly bonded carbon
n=	Non aromatic end doubly bonded nitrogen
h	Hydrogen bonded to a carbon

Table 3.3 – Table describing the FF atom type labels.

3.5. Generating a PES

3.5.1. Method

The PES displayed the relationship between the energy of a molecule and its geometry. By altering the geometry of the molecule from the ground state, the energy was increased, by plotting the energy increase as a function of distance from the ground state, generated the PES. This was discussed in more detail in Section 2.3.1. If the alterations were restricted to individual bonds, angles and dihedrals, they can they be correlated to the potential parameter sets mentioned earlier, and thus a PES can be generated for each of these. To include these potential parameter sets in the FF, the intramolecular degrees of freedom were fitted to DFT calculations, obtained using the B3LYP[14,15] functional with a 6-31G(d,p) basis set within the Gaussian03 package.[16] These generally give improved results for aromatics when compared to pure DFT methods, and because of this B3LYP was used within this work. The FF parameterisation was carried out with FF energies calculated using a specially-adapted version of the ZEBEDDE code.[13] This allowed the missing parameters in the FF to be identified, and provided calculated intramolecular energies at specific geometries for the fitting process, based on the parameters supplied. The accuracy of the energy calculation was cross checked against the Discover module of Materials Studio 5.0,[17] which was also used for the FF based geometry optimisations reported.

3.5.2. The Fitting Procedure

An example of the parameter fitting process for the cp..cp..c=1 angle potential was shown in Figure 3.5a; this corresponds to the angle between two aromatic carbon atoms and the sp^2 hybridised carbon of the imine group. This imine functionality directly attached to an aromatic ring was not specifically parameterised, and therefore did not accurately reproduce the angle upon minimisation with PCFF.

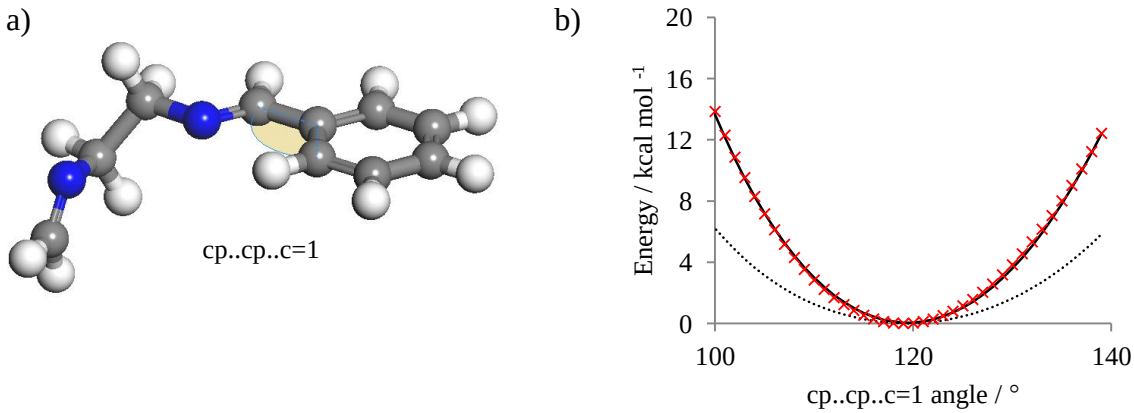


Figure 3.5 – Example of the fitting approach for intra-molecular potentials. *a)* The cp..cp..c=1 angle was shown as a shaded yellow arc, this angle was scanned at the B3LYP/6-31G(d,p) level to generate reference data. *b*) Calculated energy as a function of angle for B3LYP/6-31G(d,p) (red crosses), unmodified PCFF (dotted line), and CSFF (solid line).

Using the fragment geometry shown in Figure 3.4, a scan from 100° to 136° was carried out in 1° intervals; at each step the structure was geometry optimised with a B3LYP[14]/6-31G(d,p) level of theory, under the constraint of the fixed angle. This covered both sides of the expected minima for this angle, around 120°, therefore a PES representation could be generated. This enabled a FF parameter to be fitted to this data, with the view to successfully describe this specific angle.

A scan of this type will not only alter the angle for which parameters were required, but will also affect other geometric factors that influence the energy calculated by the FF. There will also be a knock-on effect other parameters – this was dependent on which dependency tree the *cp..cp..c=1* angle was in. To quantify this, FF atom types were assigned to the B3LYP/6-31G(d,p) optimised structures using the Discover module in Materials Studio.[17] The difference in energy from this reference for each scan point then gave the "B3LYP" and "FF" energies for comparison in the fitting procedure. The results were compared for the *cp..cp..c=1* scan in Figure 3.5b, in this case the FF underestimates the energy seen in the B3LYP data although the positions of the minimum were in good agreement. The difference between the B3LYP energy and the PCFF calculations for each scan point allowed the variation in the energy due to the missing potentials to be estimated.

The required minimum angle, θ_0, and the quadratic term for the angle potential were initially set from a quadratic fit to this difference data in the vicinity of the minimum. The full set of force constant parameters were then obtained from the full data set using linear regression with the deviation of the angle from θ_0 as the

independent variable. This example was complicated by the fact that two *cp..cp..c=1* angles were varied simultaneously in the scan, since the second exo-ring angle at the substituted cp atom has the same atom types. As a first approximation, the fitted force constants were simply halved.

This process was repeated for the entire set of intramolecular bonds, angles and dihedral potentials for the fragment. The parameters were then added to the FF and this was used to recalculate the FF estimate for the conformational energies for each of the corresponding scans. Since the energy calculated for each geometry was dependent on all the FF parameters, plotting of curves to compare B3LYP/6-31G(d,p) and FF energies for each scan showed some disagreement. Further refinement of the parameters was undertaken by cycling between fitting for individual curves, and regenerating the entire set of scan data. To compare parameter sets, we used the sum of squares:

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}\left(E_i^{B3LYP} - E_i^{FF}\right)^2 \qquad \text{(Eq. 3.1)}$$

where *i* was an index for the *N* scan conformations, and the energies, E_i, were taken from the methods indicated by the superscript labels. The process of refining the potentials was carried out with the aim of minimising this sum of squares and continued until a self-consistent parameter set was obtained. A plot including the resulting data for the example *cp..cp..c=1* angle potential was included in Figure 3.5b and it is clear that the agreement with the B3LYP reference data was very good. A full list of the parameters can be found in Appendix A.2, A.3, and A.4.

Some of the potentials fitted in this way were already present in the PCFF parameter set, most notably the aromatic ring carbons. Figure 3.6 gives a comparison of the new CSFF *cp..cp* bond, *cp..cp..cp..cp* dihedral potentials, and the original PCFF forms. The plot in Figure 3.6a shows that the *cp..cp* bond distance was shorter in CSFF than in the PCFF version by 0.033 Å (1.384 *cf.* 1.417 Å). Earlier DFT calculations using a B3LYP density functional and the TZ2P basis set on benzene[18] gave 1.403 Å and 1.391 Å, while our method gives a C-C distance in benzene of 1.397 Å. The slight shortening of the aromatic C-C distance is also consistent with the available X-ray diffraction data on the cage structures, for example the average aromatic C-C distance in **CC1a** is 1.400 Å.[1] Near to their respective minima, the PES of the *cp..cp* bond for PCFF and CSFF potentials was similar, the width of the potential well at 2 kcal mol^{-1} being 0.133

Å (CSFF) compared to 0.130 Å (PCFF), so that near the optimal bond lengths the two FF have a similar bond stiffness.

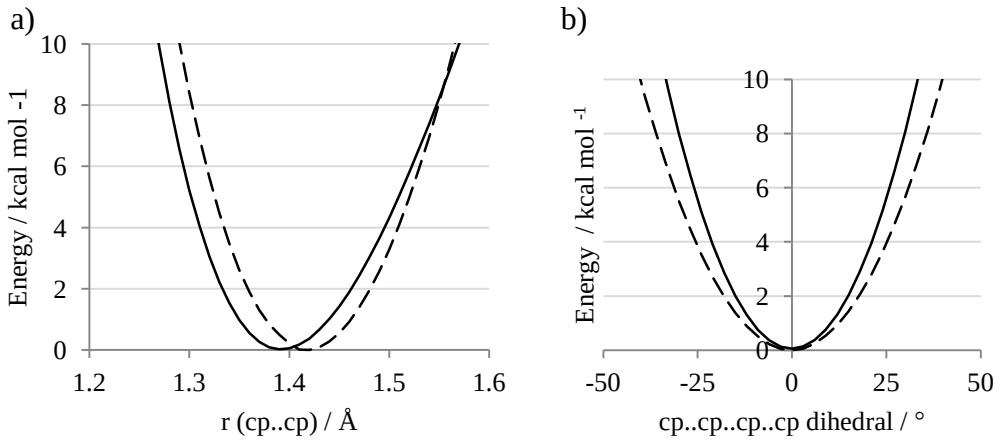


Figure 3.6 – Comparison of CSFF (solid lines) and PCFF (dashed lines) terms for aromatic carbon atoms. *a*) *cp..cp* bond stretch and *b*) *cp..cp..cp..cp* dihedral potential.

The *cp..cp..cp..cp* dihedral potential shown in Figure 3.6b shows a stronger dependence on the dihedral angle for the CSFF parameterisation than for the PCFF case. This was one of the contributions that will take part in maintaining the planarity of the conjugated aromatic and imine part of the cage structures and this difference suggested that the CSFF potential would hold in plane more closely than PCFF.

Figure 3.7 shows the dihedral potentials produced for the torsion between the imine group, the phenyl ring, and for the imine bond itself. Figure 3.7a and 3.7b demonstrate that the torsion potentials responsible for the planarity of the phenyl-imine system were basically absent from the PCFF parameterisation. For the B3LYP/6-31G(d,p) and CSFF representations, the rotation by 100° around the *cp..cp..c=1..n*= torsion has an associated barrier of around 8 kcal mol^{-1}, this lends some flexibility to the phenyl/imine system, but was still twice as large as the barrier for the unhindered rotation of the sp^3 carbon bond, exemplified by *n*=*..c..c..n*= (refer to Appendix A.4.). Comparison of Figures 3.7a and 3.7b shows that the imine bond was parameterised to give a greater energetic penalty for a given dihedral rotation than were the bonds in the aromatic ring. This was consistent with the imine having a greater double bond character than the aromatic *cp..cp* bonds.

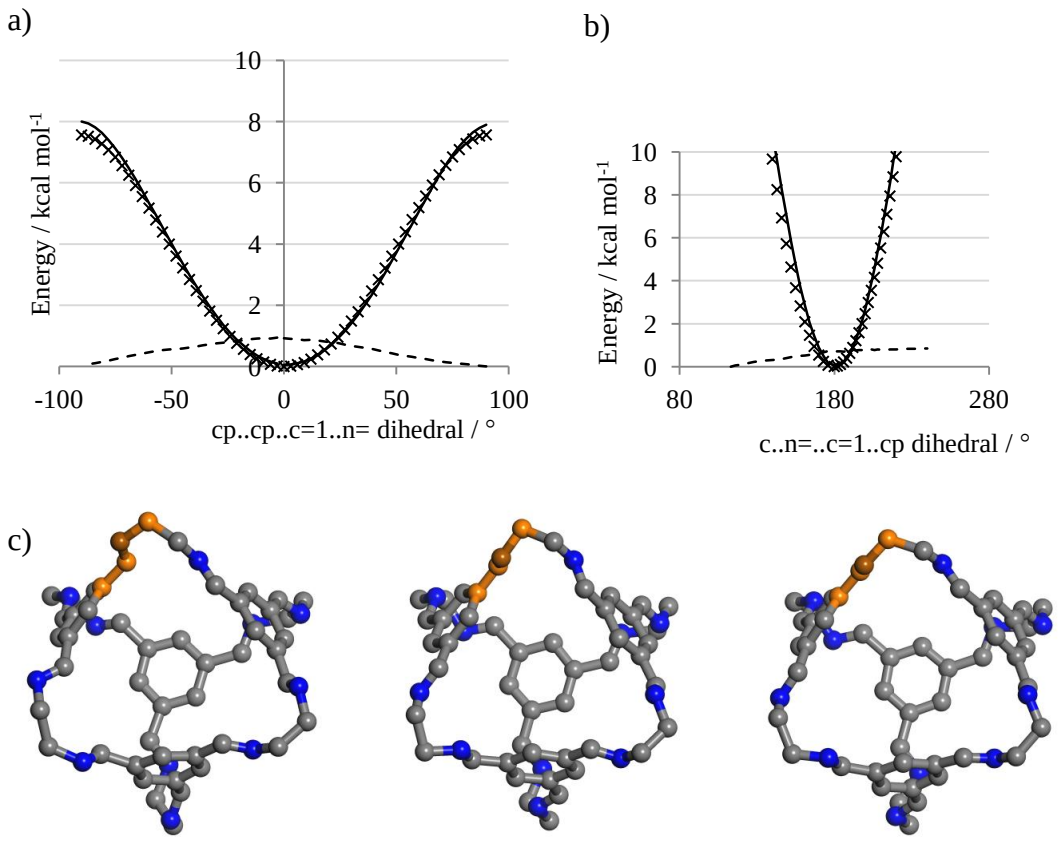


Figure 3.7 – Comparison of CSFF potential (solid lines), the PCFF potential (dotted lines) and the B3LYP/6-31G(d,p) fragment reference data (crosses) for the dihedral potentials around the imine bond. *a*) *cp..cp..c=1..n=* dihedral between phenyl ring and imine and *b*) *c..n=..c=1..cp* dihedral around the imine bond itself. *c*) Comparison of the **CC1** cage structure as optimised using PCFF and CSFF potentials and the reference crystal structure. In each case an example *cp..cp..c=1..n=* dihedral was highlighted in orange. Hydrogen atoms have been omitted for clarity.

3.5.3. Dependency trees

The bonds were the first of the potential parameter sets to be fitted in the FF, as these had the largest contribution to the overall energy. The angles were next, followed by the dihedrals. Each parameter was initially fitted in isolation to give an estimate of the final parameters. Subsequently the global implications of altering a parameter were analysed by regenerating the PES, as simulated by the FF, for every parameter, and for each iterative alteration of the specific parameter. This was necessary, as when one of the variables within a specific parameter was altered, it has a knock-on effect on any related parameters. Therefore an assessment of the knock-on effect for all the other parameters when altering a specific parameter was vital. A dependency tree shows which parameters were linked to one another; this enabled us to see what affect changing one parameter would have on any dependent parameters. Dependent parameters are any strings of FF atom types that can be found in more than one potential parameter set. By altering any of these has a direct impact any linked

potential parameter sets. For example, the *cp..cp..c=1..n=* dihedral was directly linked to both the *cp..cp..c=1* and *cp..c=1..n=* angle parameters, as well three bond parameters; *cp..cp*, *cp..c=1* and *c=1..n=*.

The RMS deviation of all the parameters was a good way to compare the FF and DFT values. Table 3.3 gives a comparison of the RMS deviation from the B3LYP/6-31G(d,p) reference data on the fragment for CSFF and for the original base FF, PCFF for each of the potentials fitted. The RMS deviation was defined using the root of equation 3.2 and can be expressed as:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(E_i^{B3LYP} - E_i^{FF}\right)^2} \qquad \text{(Eq. 3.2)}$$

This allowed the FF to be fitted, making it as accurate as possible. The CSFF better represents the DFT PES for the fragment compared to the original PCFF.

Potential	RMS deviation (kcal mol^{-1})		Potential	RMS deviation (kcal mol^{-1})		Potential	RMS deviation (kcal mol^{-1})	
	PCFF	CSFF		PCFF	CSFF		PCFF	CSFF
cp..cp	5.22	0.33	cp..cp..h	3.40	0.15	h..cp..cp..h	3.07	0.50
cp..h	3.80	0.26	cp..cp..cp	3.18	0.14	cp..cp..cp..h	3.80	0.79
cp..c=1	4.93	0.27	c=..cp..cp	2.58	0.15	cp..cp..cp..cp	9.69	2.30
c=1..h	3.48	0.22	cp..c=..h	2.49	0.09	c=1..cp..cp..cp	4.86	1.29
c=1..n=	13.95	0.41	cp..c=..n=	3.50	0.14	c=1..cp..cp..h	3.10	0.38
c..n=	4.32	0.10	h..c=..n=	2.63	0.12	h..c=1..cp..cp	2.96	0.15
c..h	3.46	0.20	c=..n=..c	2.63	0.18	cp..cp..c=1..n=	2.99	0.17
c..c	4.54	0.38	c..c..n=	4.21	0.15	c..n=..c=1..cp	9.09	0.82
			h..c..n=	5.13	0.08	c=1..n=..c..h	3.18	0.38
			c..c..h	3.98	0.14	c=1..n=..c..c	2.97	0.39
						n=..c..c..n=	3.55	0.38
						h..c..c..n=	3.58	0.34
						h..c..c..h	3.63	0.32
						c..n=..c=1..h	9.16	0.72

Table 3.3 – Comparison or the RMS deviation for all parameters for PCFF and CSFF.

3.6. Gas phase molecule testing

Now that the cage fragment has been successfully described, we can test how well CSFF described a gas phase cage. **CC1** was used for this set of calculations, since its core structure was common to all of the cage molecules (**CC1**–**CC4**). Additionally, all cage structures have the same carbon atom types in the *n=..c..c..n=* link between the imine groups (**CC1**: ethylene, **CC2**: methylethylene and **CC3**: 1,2-cyclohexyl). Figure 3.8 shows a comparison for the minimisations

of the original base FF, PCFF, the new updated FF, CSFF, and that of a cage taken from the crystal structure. It was clear that the planarity of the cage faces was more accurately reproduced using CSFF, and this was because the potential parameters required to describe the planarity of the cages was absent in PCFF. As CSFF seemed to be performing well, **CC1**, **CC2** and **CC3** were compared to higher level calculations, to see whether the FF was transferable. Each cage was optimised using DFT (B3LYP/6-31G(d,p)) within the Gaussian03 package.[16] Figure 3.8 also displays the overlay of the three cage structures obtained from the crystal structure with the structures obtained via optimisation using DFT and optimisation when using CSFF. It was evident that CSFF represents the cage accurately and this superimposition of the three indicates that the cage was well represented. This gave a molecular similarity greater than 96% based on the Field method, with an average RMS atom distance of 0.2 Å. This was similar to the B3LYP/6-31G(d,p) optimised structures that had an average molecular similarity of 98% and an average RMS distance of 0.1 Å. So although the averaged DFT structures were slightly more accurate, CSFF allows the use of molecular mechanical MD and this was computationally far less expensive than *ab initio* MD, thus making the use of CSFF far more efficient when looking at a cage in the gas phase.

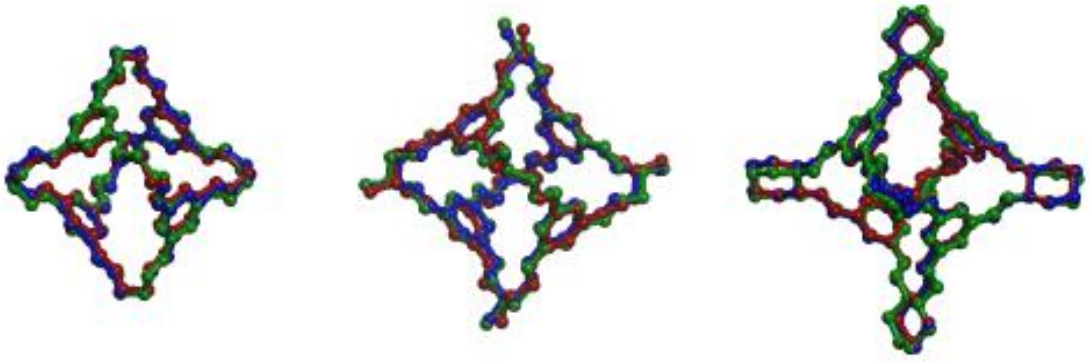


Figure 3.8 – Figure comparing the three cages, **CC1**, **CC2** and **CC3** (left to right), when taken straight from the crystallographic X-ray structure (red), minimised with CSFF (blue), minimised using DFT B3LYP calculations with a 6-31G(d,p) basis set (green) superimposed on top of one another. On average, the cages minimised using the higher level density functional theory had a 98% similarity, whereas when the cages were minimised using CSFF there was a greater than 97% similarity to the crystal data. The average RMS atom distance was 0.1 Å and 0.2 Å respectively.

As molecular flexibility was a key factor and might influence host-guest behaviour as well as gas sorption properties, it was important to make sure that CSFF was describing the flexibility of the cages reasonably. From a design perspective, these ultramicroporous cages must strike a balance between being flexible enough to allow guest movement within the voids while being rigid enough to be

shape persistent, and therefore not collapse to a denser form. Empirically speaking, the imine cages were clearly less rigid than comparably-sized molecules such as C_{60}, which can be treated as a rigid unit in MD simulations,[19] but more rigid than many organic polymer chains.[20] Therefore, the dynamic characteristics of the cages were analysed, paying particular attention to the molecular flexibility. The FF has been fitted to potential curves opposed to basic minima, consequently it would be expected to be able to perform the dynamics successfully, as the potential can oscillate at the bottom of the potential energy well.

Gas phase MD simulations were carried out for **CC1**, **CC2** and **CC3** using the Discover module in Materials Studio[17] for 1 ns using an NVT ensemble at 298 K sampled every 0.5 ps using the cage structure from the respective crystal structure model. For simplicity, only one positional isomer of **CC2** was used; an explanation for this can be found in Appendix A.1.[1] In reality, the crystal structure for **CC2** contained greater disorder due to positional isomers based on placement of the vertex methyl groups. A number of structural parameters for these cage molecules were monitored during the MD simulations, specifically around the diamine vertices: these include the C-N-C angle, the N-C-C angle, and the N-C-C-N torsion angle. These parameters were then compared to the experimental crystallographic data.[1] Since the vertices for **CC2** are asymmetric, the angle parameters behave slightly differently and hence these have been separated into C-N-C(Me), C-N-C, N-C-C(Me) and N-C(Me)-C.

When analysing the gas phase MD results, data taken directly from crystal structures may not act as a perfect benchmark as the packing forces were no longer present. However, all values seem to be in reasonable agreement at 298 K. Therefore it was unsurprising that the averages seen in Figure 3.9 deviate slightly. It was prudent to compare DFT to the FF to see how accurately the shape of the cage was being maintained. It was evident that for all three cages, the shape was being upheld. Only one conformer was considered and no interconversion was observed.[21]

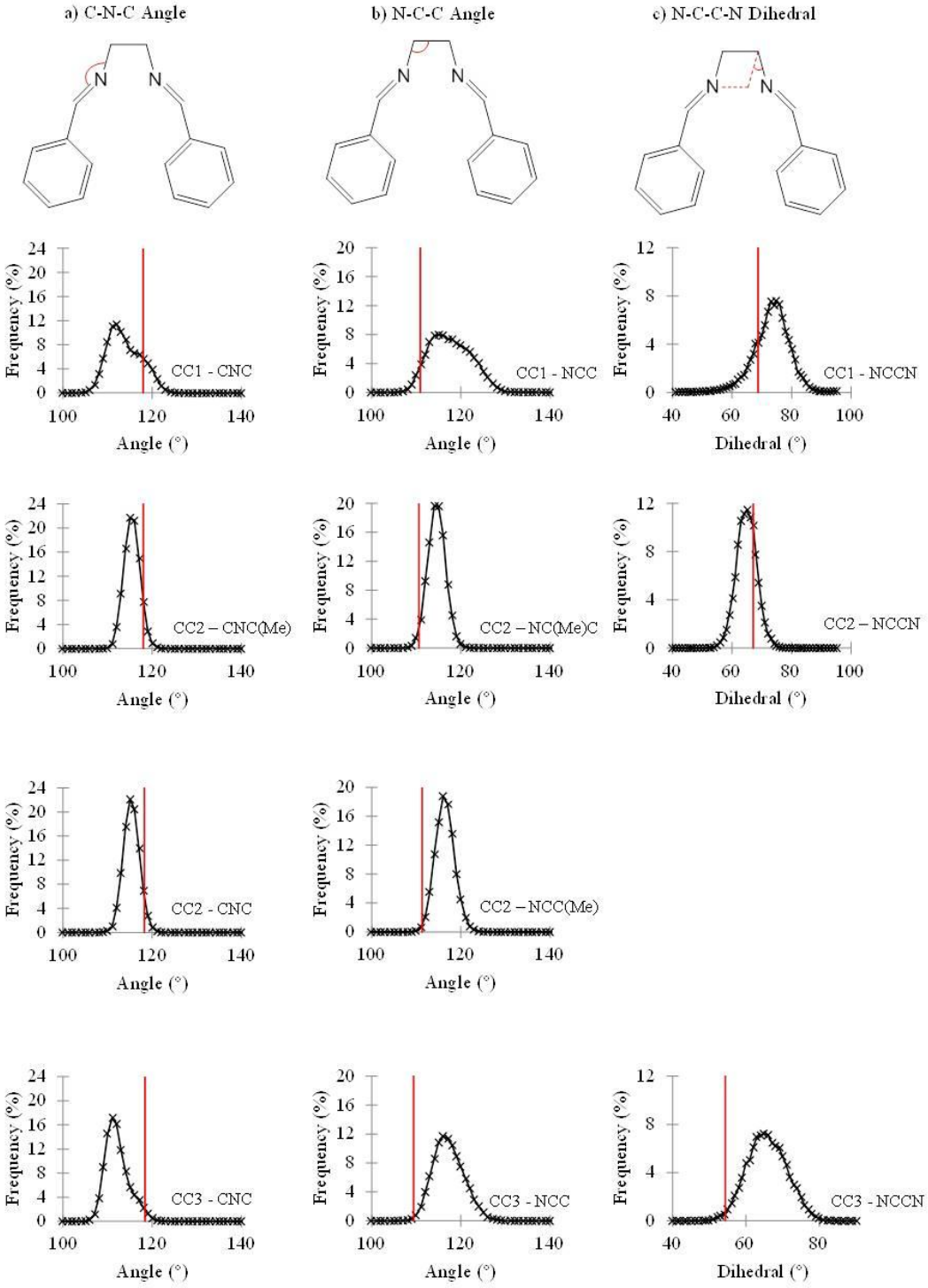


Figure 3.9 – Gas phase MD results showing a) **CC1** C-N-C angle, **CC2** C-N-C(Me) angle, **CC2** C-N-C angle, and **CC3** C-N-C angle, b) **CC1** N-C-C angle, **CC2** N-C-(Me)-C angle, **CC2** N-C-C(Me) angle and **CC3** N-C-C angle and c) the N-C-C-N dihedral for **CC1**, **CC2** and **CC3**. The red line indicates the static crystallographic value. Simulation proceeded for 1 ns using an NVT ensemble at 298 K sampled every 0.5 ps.

3.7. Testing on supercells

3.7.1. Fitting the intermolecular potentials

When initially fitting the new potential parameter sets for CSFF, only the intramolecular forces were considered. This meant that the intermolecular forces were not taken into account. Therefore it was important to see how CSFF represents the cage packing forces. This can be achieved by looking at how both optimisation and MD simulations describe the cage molecules. Before this though, it was important to look at the intermolecular potential parameter sets already present in CSFF; these will be the unmodified PCFF parameters.

3.7.2. Partial charges

These cage molecules consist of only light elements – carbon, nitrogen and hydrogen. Unlike metal-containing porous systems, the composition of the cage systems means that the charges may be less important as they are small; the maximum partial charge was located on nitrogen atom (-0.60). The contribution of the columbic interactions account for around 90% of the electrostatic interactions.[22] As PCFF has already been parameterised for organic molecules, the electrostatic interactions between atoms defined by the PCFF partial charge parameters were used without further alteration; these can be found in Table 3.4.[10]

I	J	Delta IJ	Delta JI
c	c	0.0000	0.0000
c	c=1	0.0000	0.0000
c	cp	0.0000	0.0000
c	h	-0.0530	0.0530
c	n=	0.3000	-0.3000
c=1	cp	0.0000	0.0000
c=1	h	-0.1268	0.1268
c=1	n=	0.3000	-0.3000
cp	cp	0.0000	0.0000
cp	h	-0.1268	0.1268
cp	n=	0.1993	-0.1993

Table 3.4 – Table showing the partial charge differences between atoms I and J. These were used to assign the partial charges to the different cage systems.

The partial charges were based on the following equation, where for atom *i*, the partial charge is the sum of all the charge bond increments, δ_{ij}, that:

$$q_i = \sum_j \partial_{ij} \qquad \text{(Eq. 3.3)}$$

Here *j* is all the atoms that were covalently bonded to atom *i*.

3.7.3. Intermolecular dispersion parameters

Accurate parameters were required to describe the non-bonding interactions between the individual cages. Following the PCFF approach, Lennard-Jones (LJ) 9-6 potentials were used to model repulsion-dispersion interactions:

$$E_{ij}^{LJ} = \frac{A_{\iota\kappa}}{r_{ij}^{9}} - \frac{C_{\iota\kappa}}{r_{ij}^{6}} \qquad \text{(Eq. 3.4)}$$

where the Lennard-Jones contribution to the energy for atoms *i* and *j*, which are of types ι and κ, when they were a distance r_{ij} apart. The Lennard-Jones parameters $A_{\iota\kappa}$ and $C_{\iota\kappa}$ were the potential parameters based on atom types. PCFF was parameterised using the alternative setting of Equation 3.3 that was in terms of well depth $\varepsilon_{\iota\kappa}$ and minimum position, $r_{0\iota\kappa}$. Parameters were defined for like interactions; well depths, $\varepsilon_{\iota\iota}$ and minimum positions, $r_{0\iota\iota}$, and then sixth power combining rules were used for interactions between pairs of atoms of different types. In this work we have chosen to transform the interaction parameters into the $A_{\iota\kappa}$ and $C_{\iota\kappa}$ form since this allows fitting to the X-ray crystal structure volumes of the cage structures by scaling the dispersion parameter only.

Originally, PCFF was parameterised for only "strong" intermolecular interactions, such as hydrogen bonds. This meant that molecular crystals consisting solely of "weak" van der Waals interactions were often poorly represented during a MD simulation. Within our cage systems, no such "strong" intermolecular bonds were present, and therefore the intermolecular forces need re-parameterisation. To test the importance of the absence of fitted "weak" intermolecular forces MD was used on two control systems. An aniline crystal was chosen to monitor how PCFF represented "strong" intermolecular interactions, due to the hydrogen bonds present, and a benzene crystal to see how PCFF represented "weaker" interactions. A supercell of each crystal was generated (Table 3.2), and a 20 ps NPT ensemble MD simulation, at room temperature and pressure, was simulated so that the systems had reached equilibrium. The volume remained constant for all systems post equilibration, with a fluctuation of less than 5%. Initial findings showed that the combination of the new CSFF intramolecular potential parameters and the PCFF intermolecular potential parameters consistently overestimated the unit cell volumes of these solids; this was displayed in Figure 3.10. In particular, there was a fast increase in volume for the benzene crystal over the first 3 ps of the simulation, suggesting that the PCFF parameters for the

soft interactions were too weak. Interestingly, the volume of the aniline crystal also gradually increases during the simulation, suggesting that the initial parameterisation of the intermolecular potential parameter sets in PCFF were not accurate to start with. This was especially evident for the Lennard-Jones parameters, suggesting that the attractive interaction between the cages was too weak, subsequently meaning that the cell tends to expand in the NPT ensemble simulations. This behaviour has been observed in earlier simulations of organic crystal structures. For example Pan and Chapuis reported overestimation of cell parameters in their simulation of the hexamethylenetetramine suberate structure.[23] They decided to correct the cell dimensions by imposing an external pressure tensor. Rather than take this route of changing the effective pressure of the simulation, we decided to examine the effect of scaling the dispersion coefficient, C_{IK}, within the LJ 9-6 potential on the calculated cell volume.

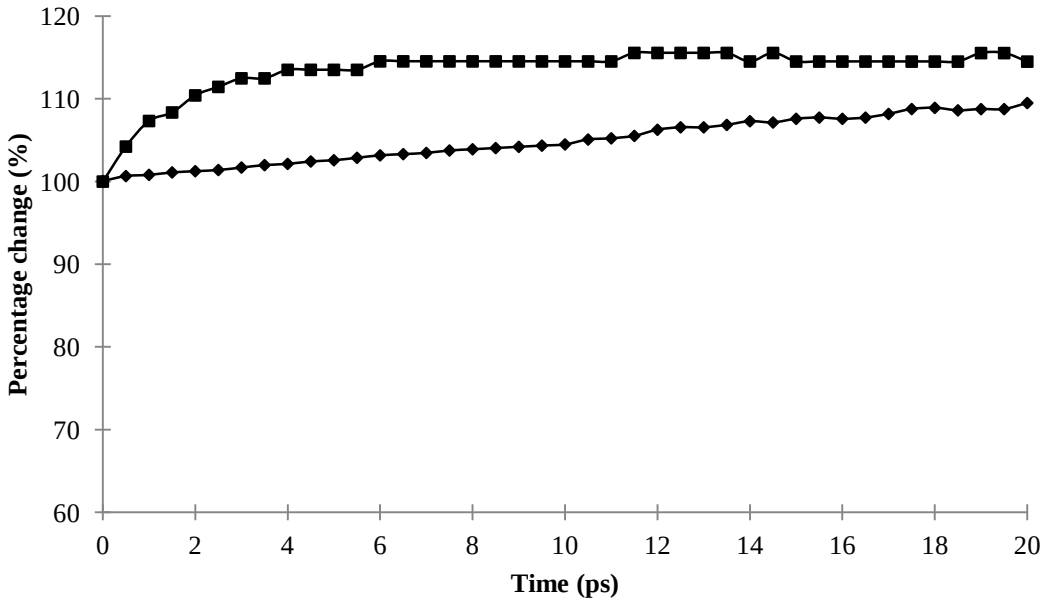


Figure 3.10 – The percentage volume (relative to the experimental starting structure) as a function of simulation time for benzene using PCFF (filled squares) and for aniline using PCFF (filled diamonds).

To scale the dispersion the C_{IK} parameter needs to be altered; the derivation for this was shown in Section 2.3.8. The original van der Waals potentials were extracted for PCFF and the existing r_o and E values were used to calculate the corresponding A_{IK} and C_{IK} parameters. The C_{IK} parameter can then be scaled as required; once the scale was decided upon, r_0 and E were calculated and these new values could be written into the PCFF file.

3.7.4. Scaling the dispersion

The influence of the intermolecular interactions on the calculated cell volume were examined using NPT MD simulations for the crystal structures of **CC1α**, **CC1β**, **CC2** and **CC3**; a simulation time of 40 ps was sufficient to reach equilibrium cell volumes in each case. These were all carried out at 298 K. The average volumes of the equilibrated systems were then compared to those observed experimentally by single-crystal X-ray diffraction. This was approached by scaling the following FF atom types within the FF by the same factor: *cp, c=1, c, n=*. It was decided that hydrogen was too light to have any significant effect on the dispersion, therefore the intermolecular potential parameter for *h* remained unaffected. The scaling of the five FF atom types and their explicit values was shown in Table 3.5. These values were input into the intermolecular potential parameter section of the adapted FF, so that everything else in the FF was consistent. This meant that the effect of the scaling could be monitored.

The intermolecular potential parameters were scaled by -10% to +90% by 10% intervals, and the above NPT MD simulations repeated. This allowed the cell volume deviation (compared to experiment), ΔV_c, to be plotted against dispersion-parameter scaling factor, Figure 3.11. All four cage systems showed a near-linear dependence of cell volume on the dispersion parameter scaling factor indicating that as the dispersion parameter was increased the calculated cell volume was reduced. The experimental volumes were reproduced (±3 %) for **CC1**–**CC3** when a scaling factor of +20 % was implemented in CSFF, and this scaling factor for the potential parameters was therefore incorporated into the FF. The effects on aniline and benzene are shown in Figure 3.12.

FF atom type	Partial Charge	r_0 (Å)	E (kcal mol^{-1})
c	0.194	3.774	0.093
cp_1	0.000	3.774	0.111
cp_2	-0.127	3.774	0.111
c=1	0.173	3.774	0.111
n=	-0.600	3.576	0.138
h_1	0.053	2.995	0.020
h_2	0.127	2.995	0.020

Table 3.5 – Table showing the partial charge, r_0 and E values for each FF atom type.

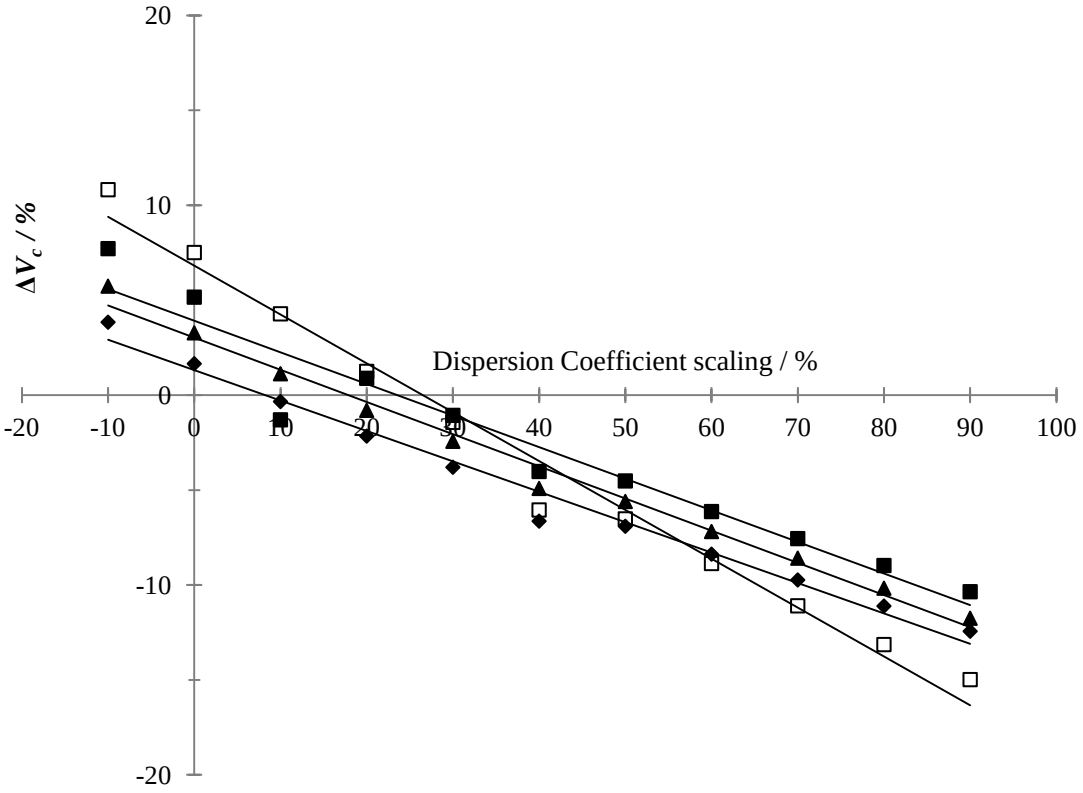


Figure 3.11 – Deviation of calculated cell volume, ΔV_c, *vs.* dispersion coefficient scaling factor for **CC1α** (open squares), **CC1β** (filled squares), **CC2** (diamonds) and **CC3** (triangles). These values were calculated using MD and were taken from the time averaged data of the equilibrated system; equilibration period of 10 ps, data collected over the following 1 ns.

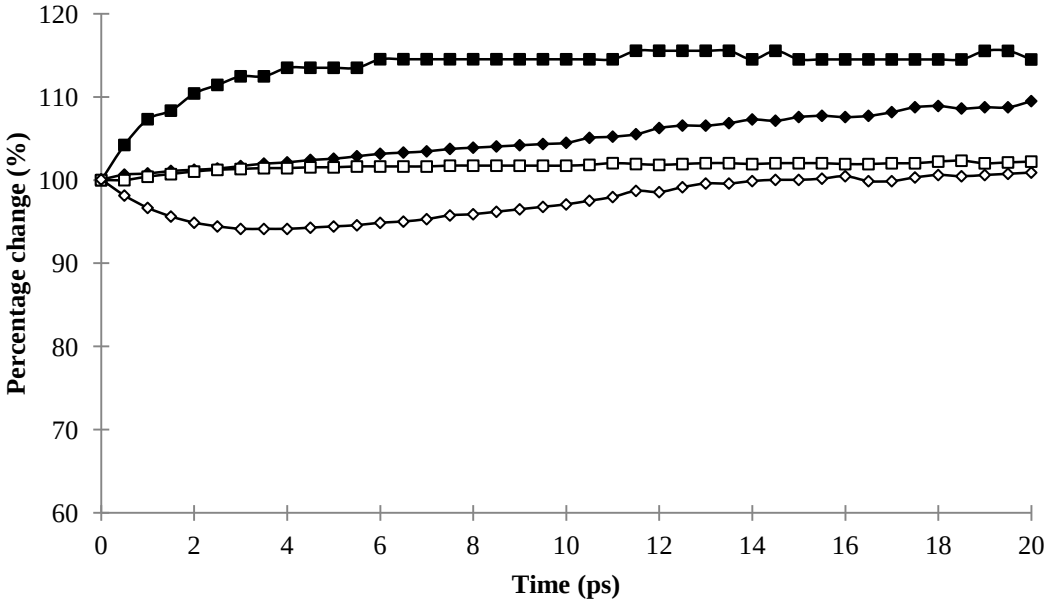


Figure 3.12 – The percentage volume (relative to the experimental starting structure) as a function of simulation time for benzene using PCFF (filled squares) and using CSFF (open squares) and for aniline using PCFF (filled diamonds) and CSFF (open diamonds).

Once the scaling of the dispersion was complete the aniline and benzene controls representing "hard" and "soft" intermolecular bonds were then simulated using CSFF, under the same conditions. Using CSFF, it was evident that both the aniline and benzene crystals volume was better reproduced; both only had a slight deviation from the expected once the systems had reached equilibrium. This helped further validate the 20 % dispersion scaling factor used within CSFF.

One issue with scaling the FF was that it was not self-consistent, as the intramolecular potentials were affected, and therefore these needed to be re-fitted, such that the PES landscapes were described accurately. Therefore, the fitting of the intramolecular and intermolecular contributions was accomplished in an iterative process.

3.7.5. Testing the effect of CSFF on the cage systems

After defining the new intramolecular potentials required for the cage structures and adjusting the dispersion interactions, we used CSFF to study the structures of the four cage systems, **CC1α**, **CC1β**, **CC2**, and **CC3**, in more detail using the supercells defined in Table 3.2. Optimisation of the structures was carried out using the Discover package within Materials Studio.[17] For this optimisation, the Smart minimiser was used which combines the steepest descent, conjugate gradient and Newton method in a cascade. The cell was allowed to relax during optimisation until the remaining force was less than 0.1 kcal mol^{-1} $Å^{-1}$, to see how well CSFF was describing the unit cell parameters upon minimisation. The same supercells were used to represent each crystal structure. Intermolecular interactions were calculated based on a potential cut off of 10 Å, and electrostatic interactions were calculated following the Ewald summation method. MD simulations employed a time step of 0.5 fs and simulation times up to 1 ns. For the fitting of dispersion parameters, NPT (constant moles, pressure and temperature) ensemble dynamics with the Hoover barostat[24] and the Hoover thermostat[24] were used to obtain calculated unit cell volumes; both had a time constant of 0.5 ps. NPT simulations were carried out at 1 atmosphere and 173 K to match the experimental conditions at which the structures were obtained. Each system was initially equilibrated with temperature scaling every 10 timesteps for 10 ps, and subsequently allowed to proceed for another 10 ps, until a steady state was reached as judged by simulation temperature and cell volume. These were run to see whether the dispersion was being held consistently over a period of time. The NPT ensemble was used, so that the lattice parameters were not restricted like an NVT ensemble. The results for both the optimisation and MD

simulations were then compared against the crystallographic data for each system in Table 3.6.

Results showed that when the **CC1α** structure was optimised, there was a small volume contraction upon optimisation, with the greatest effect in the *a*-axis of a -1.22 % shift. For **CC1β**, the *a* and *b* directions were equivalent since the cages pack hexagonally in the *a,b* plane (see Figure 3.13). On minimisation, the hexagonal symmetry was slightly distorted since the *a*-vector (-1.49 %) contraction was around 0.5 % greater than that in the *b*-direction (-0.97 %). In contrast the NPT MD simulations give an isotropic expansion from the experimental data. A similar effect was seen for the **CC2** structure: in this case there was an hexagonal arrangement in the *ab* plane but the cages were not close packed, as seen with **CC1β**, but rather arranged to form 1-D channels in the [001] direction (refer to Figure 3.13c). This gives **CC2** the lowest density of these cage structures. **CC3** shows the smallest deviation in both the optimised (0.90 %) and MD simulations (-0.72 %). This was likely to be down to the cubic close packing of the system, with the functionalised vertices most likely holding the systems shape. Figure 3.13d shows that the differences in cell parameters between experimental data and optimisation calculations were small (less than 2 %), with the NPT MD results reproducing experiment to within 1 % for all of the structures used in the fitting procedure.

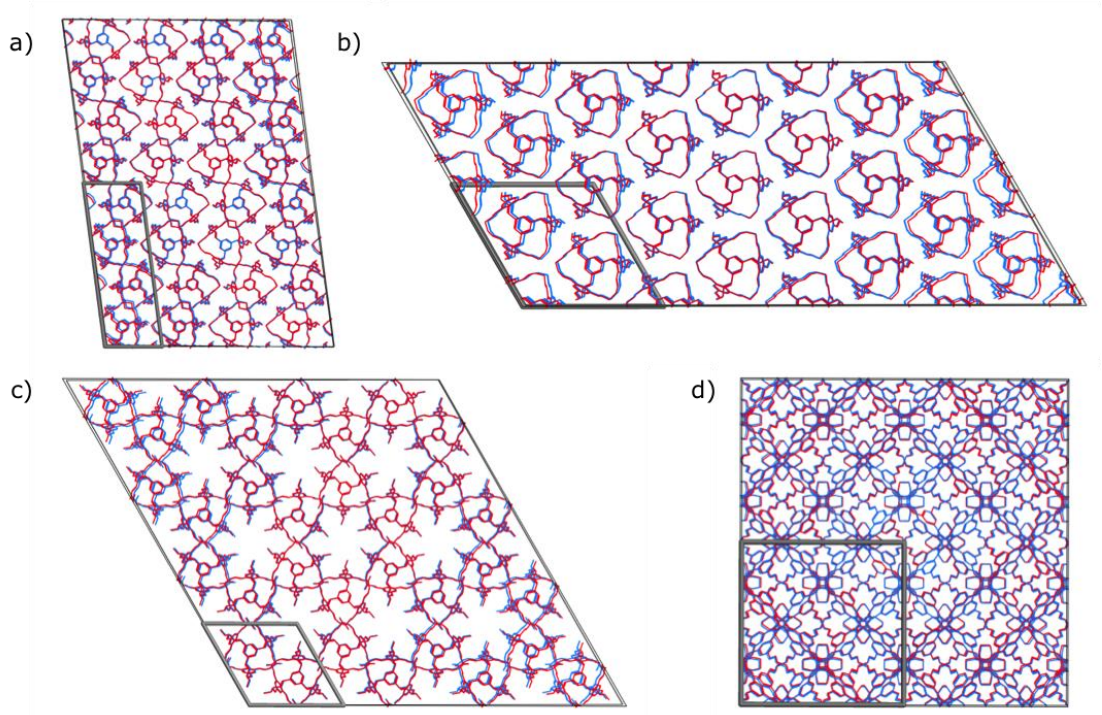


Figure 3.13 – By superimposing the CSFF minimised cage structures (blue) on top of the original X-ray crystal structure data (red), it was possible to show how well the FF represents the four systems; *a*) **CC1α** viewed down the [010] direction, *b*) **CC1β**, viewed down the [001] direction. *c*) **CC2** viewed down the

[001] direction and *d*) **CC3** viewed down the [001] direction. The grey box represents the unit cell for each cage system.

Using the MD NPT ensemble simulation, a small and isotropic expansion of the simulation cell compared with the experimental starting point was observed (+0.41%). We would expect that the MD simulation cell would have a volume greater than the minimisation result because anharmonicity in lattice vibrations will usually result in expansion of the cell dimensions compared to the "zero Kelvin" minimisation result.[25] We see from Table 3.6 that this holds for all cage systems except **CC3**.

Cage Structure	Method	Cell Parameters					
		a / Å	*b* / Å	*c* / Å	*α* / °	*β* / °	*γ* / °
1α	**CS**	**12.81**	**10.91**	**36.81**	**90.00**	**97.49**	**90.00**
4×4×2	MIN	-1.22%	-0.47%	-0.24%	0.00%	0.93%	0.00%
4×4×2	MD	0.41%	0.41%	0.41%	0.00%	0.00%	0.00%
1β	**CS**	**21.23**	**21.23**	**10.62**	**90.00**	**90.00**	**120.00**
4×2×4	MIN	-1.49%	-0.97%	-0.55%	-0.09%	-0.24%	-0.19%
4×2×4	MD	0.29%	0.29%	0.29%	0.00%	0.00%	0.00%
2	**CS**	**18.77**	**18.77**	**10.92**	**90.00**	**90.00**	**120.00**
4×4×4	MIN	-0.85%	-1.29%	-1.88%	0.12%	-0.04%	-0.41%
4×4×4	MD	-0.72%	-0.72%	-0.72%	0.00%	0.00%	0.00%
3	**CS**	**24.80**	**24.80**	**24.80**	**90.00**	**90.00**	**90.00**
2×2×2	MIN	0.90%	0.90%	0.90%	0.00%	0.00%	0.00%
2×2×2	MD	-0.27%	-0.27%	-0.27%	0.00%	0.00%	0.00%

Table 3.6 – Table comparing the percentage change of the lattice parameters from equilibrated NPT MD simulations at 173 K and cell minimisations (MIN) with the X-ray crystal structure data. Notes: *a*) The values in the first column for calculations indicate the supercell dimensions used in each case *b*) MD values were taken from the time averaged data of the equilibrated system; equilibration period of 10 ps, data collected over the following 1 ns.

3.8. MD simulations for crystalline cage systems

As for the gas phase simulations, the C-N-C angle, N-C-C angle, and N-C-C-N dihedral were monitored during MD simulations using the same conditions, and their respective ranges were again compared to crystallographic data. Figure 3.14 shows a clear agreement between the crystallographic data and the MD simulation results: the peak MD values for the various angles occur at or just off the crystallographic values. This differs from the gas phase calculations, and illustrates how important the crystal packing forces were in maintaining the cage

geometry. This also indicates that although CSFF was fitted to just a 24 atom fragment, it describes accurately the motion of the extended crystal structures.

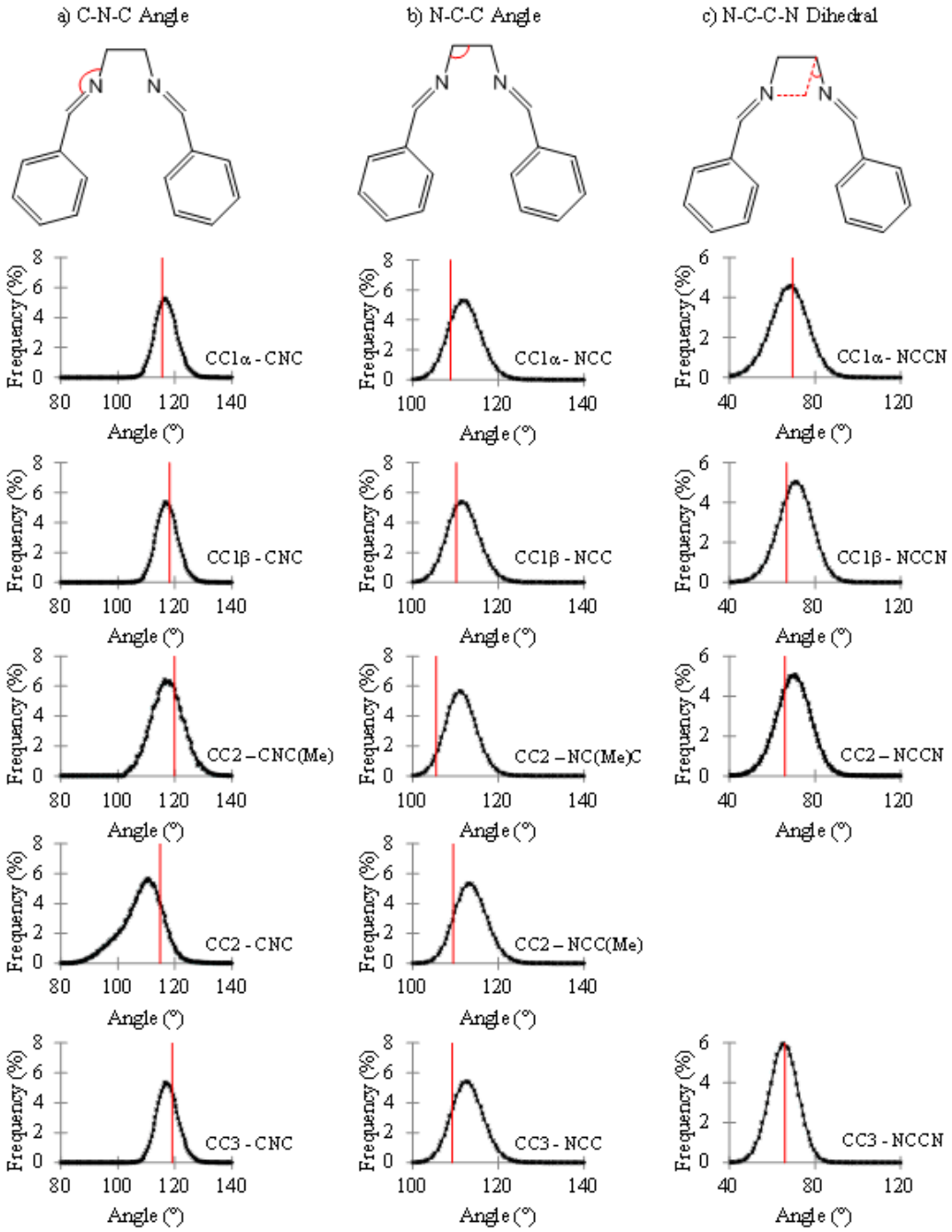


Figure 3.14 – Crystal structure MD results showing a) **CC1α** C-N-C angle, **CC1β** C-N-C angle, **CC2** C-N-C(Me) angle, **CC2** C-N-C angle, and **CC3** C-N-C angle, b) **CC1α** N-C-C angle, **CC1β** N-C-C angle, **CC2** N-C-(Me)-C angle, **CC2** N-C-C(Me) angle and **CC3** N-C-C angle. Finally c) displays the N-C-C-N dihedral for **CC1α**, **CC1β**, **CC2** and **CC3**. The red line indicates the static crystallographic value. Simulation proceeded for 1 ns using an NVT ensemble at 298 K sampled every 0.5 ps.

Looking at each structural parameter in more detail: Figure 3.14a displays the range for the angle linking the cage face to its vertices, the C-N-C angle. This was

modelled using the data from the PES generated for the cage fragment (Figure 3.4), and has a minimum of 118°. This compares well with the crystallographic values for the cages, which range between 115–119°. Figure 3.14b displays the variation in the N-C-C angle. This angle determines how spherical the cage is: if this angle was decreased excessively, then it would cause the cage to be 'squashed'. For the symmetrical cages, **CC1**α, **CC1**β and **CC3**, there was an oscillation in the distribution of angles centring on the crystallographic data, indicating that the cage shape was well maintained. **CC2** was less symmetrical in nature due to the disordered positioning of the methyl groups on the cage vertices, and the use of a single isomer for the simulations of **CC2**, as described above, might account for the slight offset of the MD values with respect to experiment.

The MD results also describe the twisting observed in the cage vertices, as shown in Figure 3.14c, based on the N-C-C-N torsional structure parameter. This was modelled on the three-term cosine curve, fitted to the PES generated previously using the cage fragment; this was shown in more detail in Appendix A.4. Again, there was very close agreement between simulation and experiment; the maxima in the MD range were close to the crystallographic data for all cages. Taken together, this correspondence between simulation and crystallography suggests that CSFF was well suited to examining how molecular flexibility influences the physical properties of these cages, such as guest uptake and, perhaps, guest selectivity.

3.9. Blind testing the FF

To check the transferability of CSFF for porous organic imine cages, two systems that were *not* used in the parameterisation were also examined. First **CC4**,[8] the cyclopentyl analogue of **CC3**, was examined. Second, we studied the co-crystal formed from an equimolar mixture of cages **CC1** and **CC3**.[3] This co-crystal was "quasi-racemic", in that it only incorporates the *S*-enantiomer of **CC1** with the chirally pure *R*-enantiomer of **CC3**. The structures were optimised using the Discover package within Materials Studio,[17] using the Smart minimiser. The cell was allowed to relax during optimisation until the remaining force was less than 0.1 kcal mol^{-1} $Å^{-1}$. Subsequently, both systems were used to simulate a 1 ns NPT MD simulation, with an equilibration time of 10 ps. The results for both the optimisation and MD simulations were then compared against the crystallographic data for each system in Table 3.7.

In both cases, Table 3.7 shows that minimisation of the structures led to an overall contraction in the cell volume with some differences between the crystallographic directions, whereas the NPT MD simulations lead to smaller, uniform contraction (**CC4**) or a small expansion (**CC1**/**CC3** co-crystal). Here the differences between the experimental and calculated lattice vectors were below 1%, with the NPT MD calculations showing the smallest differences. This was analogous to what was seen for the original set of cages, **CC1-3**. This suggests that the CSFF potentials should be able to describe cage structures containing substituted imine groups that were not present in the parameterisation data set, provided that their own functionality has been carried out. Superimposition of the crystal structure and minimised data is shown in Figure 3.15.

		Cell Parameters					
	Method	a / Å	b / Å	c / Å	α / °	β / °	Υ / °
CC4	**CS**	**47.846**	**47.846**	**43.828**	**90.00**	**90.00**	**120.00**
	MIN	-0.86%	-0.90%	-0.84%	-0.08%	-0.07%	-0.19%
	MD	-0.49%	-0.49%	-0.49%	0.00%	0.00%	0.00%
CC1α/3	**CS**	**48.548**	**48.548**	**48.548**	**90.00**	**90.00**	**90.00**
	MIN	-0.96%	-0.96%	-0.96%	0.00%	0.00%	0.00%
	MD	0.26%	0.26%	0.26%	0.00%	0.00%	0.00%

Table 3.7 – Table comparing the percentage shift of the lattice parameters post MD and minimisation when compared to the experimental X-ray crystal structure data.

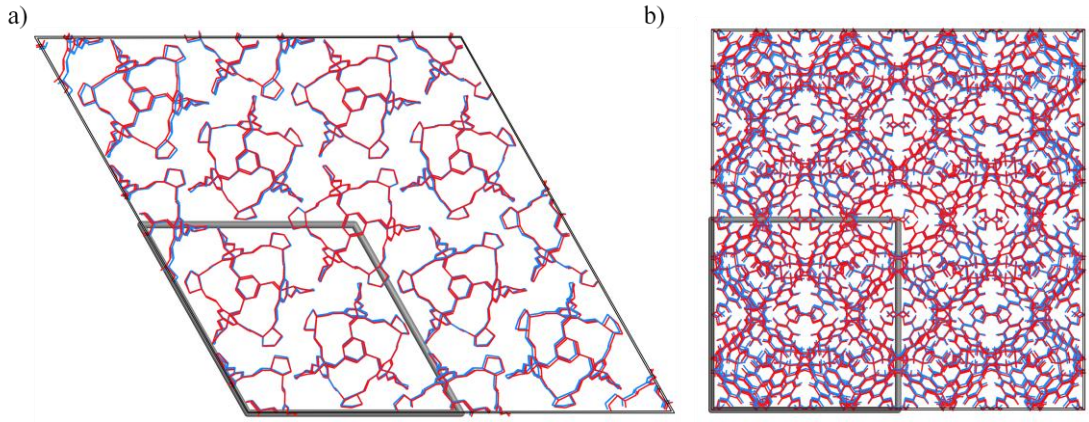


Figure 3.15 – By superimposing the minimised cage structures (blue) on top of the original X-ray crystal structure data (red), it was possible to show how well the FF represents the four systems; a) **CC4**, b) cocrystal of two separate cages (**CC1α**, **CC3**). The grey box represents the unit cell for each cage system.

3.10. Fitting of the gases

Now that CSFF has been reparameterised and subsequently validated to accurately describe the dynamics and flexibility of the cage structures, it was important to accurately describe the behaviour of any gases that were to be

modelled within the cage systems. Three classes of gases were chosen; 'linear' gases, tetrahedral or octahedral gases, and spherical gases. Within each of these classes, several small molecules were defined. Three 'linear' gases were chosen; hydrogen, nitrogen and carbon dioxide. One octahedral gas was chosen, sulfur hexafluoride, and one tetrahedral gas, methane, was used. Finally, two spherical gases were input into the FF; krypton and xenon. To define the gases, existing literature data was collated and tabulated (Table 3.8) and the following parameters used. 9-6 LJ were chosen and therefore the parameters for these were matched across against their 12-6 LJ analogues. This allows the van der Waals radii to be calculated; note this was the minimum radii of the gas and therefore should be used when considering connectivity – this does not necessarily mean that these values should be used for calculating surface areas and were only a measure of the connectivity of the system.

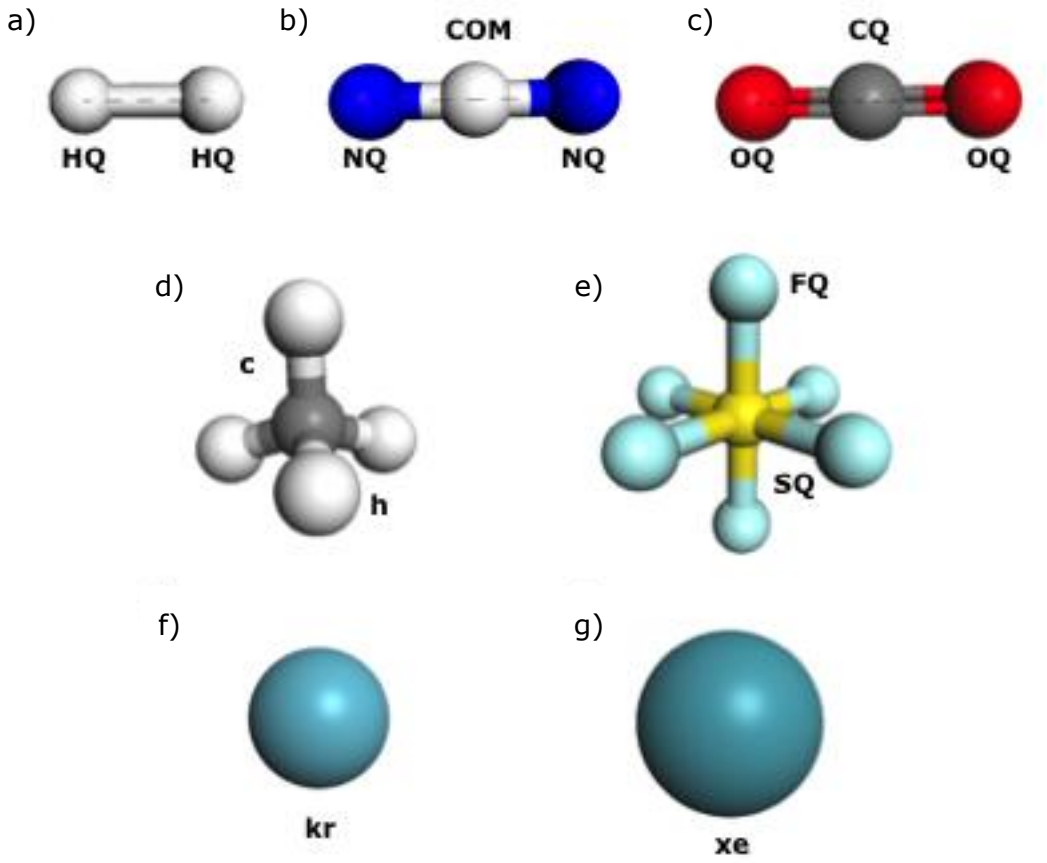


Figure 3.16 – Figure illustrating the different gases fitted within CSFF. The three linear gases were a) hydrogen, b) nitrogen and c) carbon dioxide. d) Shows the tetrahedral methane, and, e) the octahedral sulfur hexafluoride. Two noble gases were also fitted, f) krypton and g) xenon.

3.10.1. Linear Gases

Hydrogen was modelled as a diatomic model and not as a three-site multipole, as suggested in some of the literature, and can be seen in Figure 3.16a. The parameters were taken from Zecchina *et al.*;[11] here they fit a 9-6 LJ potential (which can they be inserted into PCFF) to very accurate anisotropic data from Diep and Johnson.[26] The molecule was kept rigid throughout. The FF atom type was *HQ*, with a *HQ-HQ* bond length of 0.741 Å. The van der Waals radii used for hydrogen was 1.358 Å.[27]

A three-site multipole was used to simulate nitrogen. Again, this was kept rigid and the van der Waals were based on work done previously by Potoff *et al.*[28] For this a 9-6 LJ potential was matched to the 12-6 LJ potential found in the literature. FF atom types involved were *COM* and *NQ,* where *COM* was a charged massless site. It has a *NQ-NQ* bond length of 1.10 Å and the van der Waals radius used for nitrogen was 1.55 Å.[27] Figure 3.16b shows a schematic representation of this.

Carbon dioxide was also simulated as a three-site multipole and, again, this was kept rigid, based on work by Potoff *et al.*[28] For this a 9-6 LJ potential was matched to the 12-6 LJ potential found in the literature, this was shown in Appendix A.5. FF atom types involved were *OQ* and *CQ.* Figure 3.16c illustrates a carbon dioxide molecule and this has a *CQ-OQ* bond length of 1.16 Å. The atom sizes used for *CQ* and *OQ* were 1.70 Å and 1.55 Å respectively.

3.10.2. Tetrahedral and octahedral gases

All models for methane found in literature act as a united atom model. As shape was deemed important in this work, these models were deemed unsatisfactory and therefore a five-site tetrahedral model was used.[29] This was again kept rigid and the FF was used to generate the potentials. The FF atom types involved were *c* and *h*. Figure 3.16d illustrates a methane molecule and this has a C-H bond length of 1.10 Å.

Sulphur hexafluoride was based on a seven-site octahedral complex, which was kept rigid. For this a 9-6 LJ potential was matched to the 12-6 LJ potential found in the literature.[30] The FF atom types involved were *SQ* and *FQ* and these have the size 1.80 Å and 1.47 Å respectively. An illustration of this can be found in Figure 3.16e and the *SQ-FQ* bond length was 1.564 Å. Figure 3.17 shows how the size of both these gases was calculated.

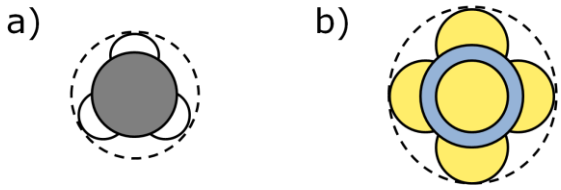


Figure 3.17 – Figure showing how the size of a) methane and b) sulfur hexafluoride were initially calculated.

3.10.3. Spherical gases

Both krypton and xenon were based on recent work by Sikora *et al*.[31] The 9-6 LJ potentials were matched to the 12-6 LJ potentials. The FF atom types were *kr* and *xe* and these have a radius of 1.818 Å and 2.05 Å, respectively, Figure 3.16f and 3.16g.

3.10.4. Parameters for the gases

The overall parameters for the gases can be seen in Table 3.8, with their associated references. The parameter curves can be located in Appendix A.5.

FF Atom Type	r_0 (Å)	ε (kcal/mol)	Charge	r, when V = 0 (Å)	Source
HQ	3.109	0.027	0.000	1.358	[11]
CQ	3.160	0.054	0.700	1.380	[28]
OQ	3.470	0.155	-0.350	1.515	[28]
NQ	3.760	0.072	-0.482	1.550	[28]
COM	-	-	0.964	-	[28]
C in CH_4	3.774	0.093	0.000	1.648	[29]
H in CH_4	2.995	0.020	0.000	1.308	[29]
Kr	4.130	0.330	0.000	1.803	[31]
Xe	4.670	0.439	0.000	2.035	[31]
SQ	3.660	0.330	0.660	1.598	[30]
FQ	3.350	0.053	-0.110	1.463	[30]

Table 3.8 - Table of 9-6 LJ parameters converted from 12-6 LJ parameters. *HQ* was already in 9-6LJ format and therefore remains the same.

3.11. Conclusion

To conclude, we have shown that it has been possible to parameterise a FF using a combination of X-ray crystal structure data with high level DFT calculations, and that this FF accurately describes the porous organic cages used in the parameterisation (**CC1**, **CC2**, **CC3**), as well as being transferable to other systems such as **CC4** and a co-crystal of two cages, (**CC1α**, **CC3**). The main adaption of the FF that was required involved parameters for the imine group conjugated to an aromatic system. It should be expected that the parameters

obtained will improve the description of similar systems when using PCFF. The scaling of the dispersion parameters to obtain reasonable crystal structures appears to be a more general requirement to use PCFF for inter-molecular van der Waals forces. In addition, seven different gases have been inserted into CSFF, so that MD simulations of gases incorporated within cage systems can be simulated. These should aid with the understanding of how and why these porous organic cage systems were important as well as providing insight into their separation capabilities, as discussed in Chapters 4 and 5.

The generation of CSFF[32] has made it feasible to simulate how the diffusion of small gases was altered dependent on their particle size. Using MD simulations, we have unlocked phenomena such as gas selectivity, rare-event hopping, and displacement of gases to regions previously considered inaccessible; all of which help to rationalise existing experimental observations. In addition, a new visualisation technique has been developed to show both the pore topology and dynamic connectivity of the system whilst a MD simulation progresses.[33] This unlocks new ideas and suggests how the pore channels of various systems previously thought too small for gas adsorption are actually ideal candidates for separation.

In the future, it will therefore be possible to use of CSFF to help rationalise the uptake of gases, and also larger guests, such as halogens and solvents, within a whole host of different cage systems. The view here was to use MD analysis for *in silico* screening of cage materials, in particular molecular separations. If reliable, this could be faster than the associated sorption experiments.

3.12. References

(1) Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. Nat. Mat. **2009**, *8*, 973–978.

(2) Hasell, T.; Chong, S. Y.; Jelfs, K. E.; Adams, D. J.; Cooper, A. I. J. Am. Chem. Soc. **2012**, *134*, 588–598.

(3) Jones, J. T. A.; Hasell, T.; Wu, X.; Bacsa, J.; Jelfs, K. E.; Schmidtmann, M.; Chong, S. Y.; Adams, D. J.; Trewin, A.; Schiffman, F.; Cora, F.; Ben Slater; Steiner, A.; Day, G. M.; Cooper, A. I. Nature. **2011**, *474*, 367–371.

(4) Sudik, A. C.; Millward, A. R.; Ockwig, N. W.; Côté, A. P.; Kim, J.; Yaghi, O. M. J. Am. Chem. Soc. **2005**, *127*, 7110–7118.

(5) Yoshizawa, M.; Klosterman, J. K.; Fujita, M. Angew. Chem. Int. Ed. **2009**, *48*, 3418–3438.

(6) Fujita, M.; Umemoto, K.; Yoshizawa, M.; Fujita, N. Chem. Commun. **2001**, 509-518.

(7) Jones, J. T. A.; Holden, D.; Mitra, T.; Hasell, T.; Adams, D. J.; Jelfs, K. E.; Trewin, A.; Willock, D. J.; Day, G. M.; Bacsa, J.; Steiner, A.; Cooper, A. I. Angew. Chem. Int. Ed. Engl. **2011**, *50*, 749–753.

(8) Mitra, T.; Wu, X.; Clowes, R.; Jones, J. T. A.; Jelfs, K. E.; Adams, D. J.; Trewin, A.; Bacsa, J.; Steiner, A.; Cooper, A. I. Chem. Eur. J. **2011**, **17**, 10235–10240.

(9) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. J. Am. Chem. Soc. **2001**, *114*, 10024–10035.

(10) Sun, H. Macromolecules. **1995**, *28*, 701-712.

(11) Vitillo, J. G. J.; Ricchiardi, G. G.; Spoto, G. G.; Zecchina, A. A. Phys. Chem. Chem. Phys. **2005**, *7*, 3948–3954.

(12) Sun, H. J. Phys. Chem. B. **1998**, *102*, 7338–7364.

(13) Lewis, D. W.; Willock, D. J.; Catlow, C. R. A.; Thomas, J. M.; Hutchings, G. J. Nature. **1996**, *382*, 604–606.

(14) Becke, A. D. J. Chem. Phys. **1993**, *98*, 5648–5652.

(15) Lee, C.; Yang, W.; Parr, R. Phys. Rev. B. **1988**, *37*, 785–789.

(16) Gaussian 09 Software Program.

(17) Accelrys; Accelrys Software, Inc.: San Diego, **2011**.

(18) Miani, A.; Cané, E.; Palmieri, P.; Trombetti, A.; Handy, N. C. J. Chem. Phys. **2000**, *112*, 248.

(19) Malaspina, T.; Fileti, E. E.; Rivelino, R. J. Phys. Chem. B. **2007**, *111*, 11935-11939.

(20) Larsen, G. S.; Lin, P.; Siperstein, F. R.; Colina, C. M. Adsorption. **2010**, *17*, 21–26.

(21) Jelfs, K. E.; Schiffmann, F.; Jones, J. T. A.; Slater, B.; Cora, F.; Cooper, A. I. Phys. Chem. Chem. Phys. **2011**, *13*, 20081–20085.

(22) Leach, A. R. *Molecular Modelling: Principles and Applications,* Prentice-Hall, **2001**.

(23) Salje, E. K. H.; Hayward, S. A.; Lee, W. T. Acta Cryst. **2004**, *61*, 3–18.

(24) Melchionna, S.; Ciccotti, G.; Lee Holian, B. Mol. Phys. **1993**, *78*, 533–544.

(25) Greathouse, J. A.; Allendorf, M. D. J. Phys. Chem. C. **2008**, *112*, 5795–5802.

(26) Diep, P.; Johnson, J. K. J. Chem. Phys. **2000**, *112*, 4465.

(27) http://www.ccdc.cam.ac.uk/products/csd/radii.

(28) Potoff, J. J.; Siepmann, J. I. AIChE J. **2004**, *47*, 1676–1682.

(29) Mesli; Mahboub; Mahboub Arab. J. Chem. **2011**, *4*, 12–12.

(30) Beniere, F. M.; Boutin, A.; Simon, J. M.; Fuchs, A. H.; de Feraudy, M. F.; Torchet, G. J. Phys. Chem. **1993**, *97*, 10472–10477.

(31) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L.; Snurr, R. Q. Chem. Sci. **2012**, *3*, 2217.

(32) Holden, D.; Jelfs, K. E.; Cooper, A. I.; Trewin, A.; Willock, D. J. J. Phys. Chem. C. **2012**, *116*, 16639–16651.

(33) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Microporous Mesoporous Mater. **2012**, *149*, 134–141.

Chapter 4

Understanding 'on'/'off' porosity switching in a molecular organic solid, CC1

4.1. Introduction

Selective adsorption of gases within porous materials is an area that has generated much interest.[1-7] This is a desirable function for biological membranes[1-7], and is also an important feature when considering systems for gas separation.[8-10] It has been shown that porous materials can respond in a structural manner towards guest sorption,[8,11-15] suggesting that they could be used to mimic enzymes.[1,3,16,17] This accommodation in structure can have implications that are not explained by single crystal X-ray data, which gives only a static representation.[15,18] Adsorption of gases can involve cooperative mechanisms. This has been seen in several porous crystals, for example calixarenes[8,9,19,20] and clathrates.[8,11-15,21] It is rare to find permanent pore structures within molecular crystals that permit sorption of gases, and one reason for this is the absence of strong intermolecular covalent or coordination bonding.[1,5] This has opened up the possibility of guest induced transformations within porous crystals, brought about by the positional translation or rotation of the discrete building blocks. This allows materials that one might expect to be non-porous, to actually be permeable to small gases, for example hydrogen and nitrogen.[2]

If discrete molecules can undergo single-crystal-to-single-crystal transformations, one might design a material that could have its porosity switched 'on' and 'off'. Previously, we showed that this is indeed possible; the porosity in **CC1** was switched 'on' and 'off' in response to a chemical stimulus.[8] This was achieved by switching between two polymorphs. To date, **CC1** has been found to exist as three polymorphs; one which is non-porous to both hydrogen and nitrogen (**CC1α**), one which is selectively porous to hydrogen over nitrogen (**CC1β**), and one which is porous to both hydrogen and nitrogen (**CC1γ**).[4,6,22] The sorption properties of **CC1** are shown in Figure 1.6. Because porosity and permeability are virtually synonymous, for a system to be porous, it must allow gas to permeate freely through the system; this shows the importance of pore connectivity.[23]

The 'chemical stimuli' for the formation of these polymorphs are the solvents ethyl acetate (**CC1α**), dichloromethane (**CC1β**), and *o*-xylene (**CC1γ**), and these are shown in Figure 4.1.[22] **CC1α** and **CC1β** can undergo interconversion *via* a single-crystal-to-single-crystal transformation, by the exposure of **CC1α** to dichloromethane, and **CC1β** by exposure to ethyl acetate. This means that the hydrogen porosity can be switched either 'on' or 'off', and that this is a reversible process; during the interconversion, the conformation of the cage changes as well

as its packing in the crystal lattice. This transition is facilitated *via* an intermediate solvated structure.[8]

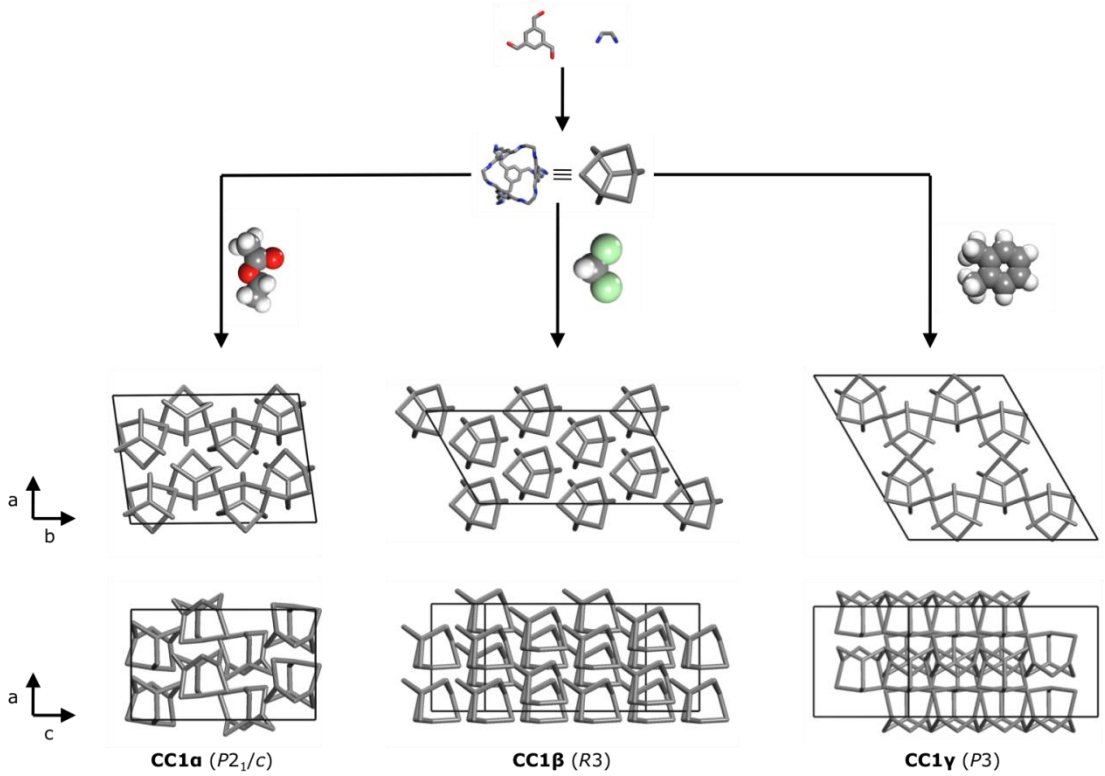


Figure 4.1 – Scheme showing the triformyl benzene and ethylene diamine building blocks of **CC1** (top) and, subsequently, how **CC1** can be crystallised into three different polymorphs. This was accomplished *via* crystallisation in the presence of different solvents; ethyl acetate (**CC1α**), dichloromethane (**CC1β**), and *o*-xylene (**CC1γ**).[22] A supercell of the single crystal X-ray diffraction structure is shown for **CC1α** and **CC1β**, while **CC1γ** is represented as an assumed structure, as no X-ray diffraction data was available.

4.2. Simulation Details

To help understand why gas uptake was observed, as well as to identify any differences between the polymorphic states, a series of simulations were undertaken. Two gases were studied in detail: hydrogen and nitrogen. These were chosen because the porosity of **CC1** to hydrogen was dependent on which polymorph is present, while **CC1** is non-porous to nitrogen for both **CC1α** and **CC1β**. It was not possible to obtain single crystal data for **CC1γ**, so X-ray powder diffraction was used to obtain the space group and lattice parameters of the system. From this, **CC1γ** was assumed to pack like **CC2**, and this was used instead to draw the schematic of the polymorph. However, the experimental sorption for **CC1γ** (Figure 4.2) showed that when **CC1** has the correct packing motif, it can be porous to nitrogen. It is thought to pack in a similar manner to **CC2**, with a two-dimensional pore channel running extrinsically through it. This would explain the hydrogen and nitrogen uptake,[24] and that the sorption

capabilities of **CC1** are dictated by its packing in the solid state, and not solely by the discrete molecular structure.

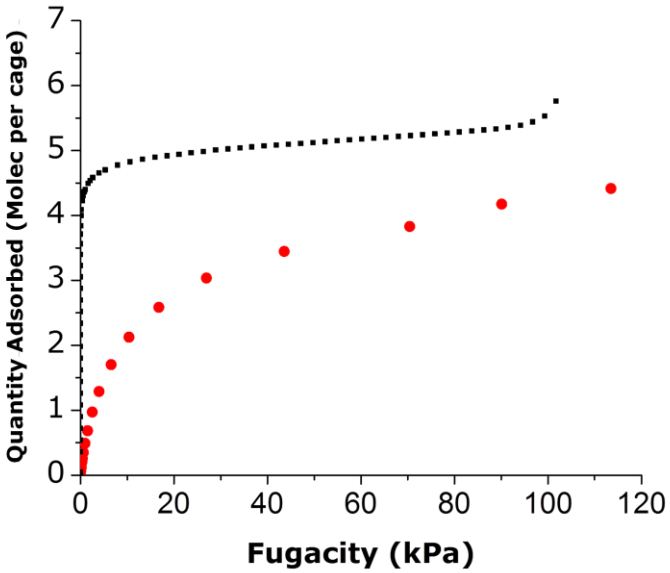


Figure 4.2 – Figure showing the experimental uptake of hydrogen (black) and nitrogen (red) for **CC1γ**.[22]

To see whether the porosity of **CC1** was dictated by the discrete unit or by its crystal packing, three types of simulations were undertaken for **CC1α** and **CC1β**:

1. Analysis of the static crystal structures to determine the connectivity towards hydrogen and nitrogen;
2. Simulated sorption properties to look at the theoretical uptakes of hydrogen and nitrogen and the corresponding loading profiles;
3. Molecular dynamics (MD) simulations to ascertain the permeability of hydrogen and nitrogen; this will help rationalise what has been shown experimentally. If applicable, diffusion coefficients can be calculated.

The accessible surface area was calculated and generated using the molecular simulation program Zeo^{++} upon the 'static' X-ray crystal structures.[11,13] The probe radius was altered so that the different pore topologies for hydrogen (1.09 Å) and nitrogen (1.55 Å) could be quantified and visualised.

The thermodynamic properties of the hydrogen and nitrogen were initially examined using the Sorption tool within Accelrys' Materials Studio[16,17] using CSFF to simulate the adsorption isotherms *via* Grand canonical Monte Carlo (GCMC) simulations. For each gas, single component isotherms were generated. The gas molecules were randomly weighted, translated and deleted with equal weighting within a 2 x 2 x 2 supercell of **CC1α** and a 2 x 2 x 1 supercell of **CC1β**, using the Metropolis Monte Carlo method, treating the cage host as a rigid body. The

fugacity range was 0.01 to 100 kPa. This was carried out at 77 K to match the experimental sorption conditions. 100,000 equilibration steps were used with a production run of 1,000,000 steps. The density profile was calculated by sampling the sorption sites every 50 steps, with a grid interval of 0.4 Å.

As **CC1** is flexible in nature, it was important to see how much this flexibility affected the connectivity of the polymorphs. Therefore, two different types of MD simulations were undertaken. One where the cages were treated as rigid bodies, the rigid body method (RBM), and one where the cage flexibility was described using CSFF; this is referred to as the flexible body method (FBM). The RBM restricted the movement of the cages within the system, treating each cage as a rigid body, meaning that only the intermolecular potentials were required; these were modelled using CSFF. This is different to fixing the framework of a porous network, since the discrete units are able to move with respect to one another. This means that any movement was achieved by the libration of the molecules against one another, and not by their internal movement. The other simulation used was the FBM, and this used CSFF to describe the cage flexibility; the parameterisation of this can be found in Chapter 3.

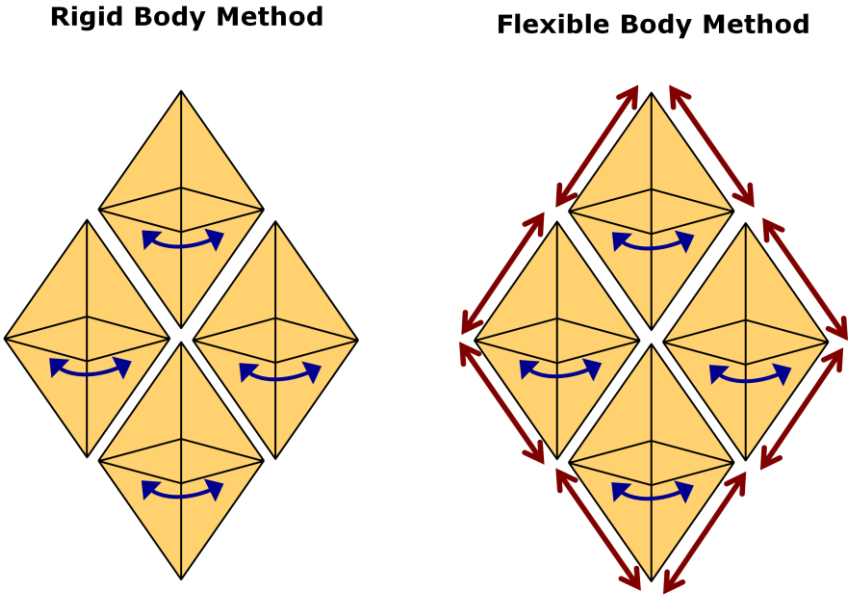


Figure 4.3 – Schematic showing the difference between the MD simulations. The blue arrows indicate the intermolecular translations and rotations of the cage bodies, whilst the red arrows indicate the introduction of the intramolecular degrees of freedom.

By comparing the differences between the MD simulations, insight into how the diffusion of gases was affected by the flexibility of the system was obtained. The differences between the simulations are shown schematically in Figure 4.3.

A 2 x 2 x 2 supercell of both polymorphs, containing several thousand atoms, was used to represent the crystal structure; these had the initial cell dimensions of **CC1α**: A = 51.24 Å, B = 43.64 Å, C = 73.62 Å, α = γ = 90°, β = 97.49° and **CC1β**: A = 84.92 Å, B = 42.46 Å, C = 42.48 Å, α = β = 90°, γ = 120°. A single molecule of gas was then inserted into each supercell, so that it was in the centre of one of the cage voids. A detailed description on how each gas was modelled can be found in Section 3.10. A van der Waal potential cut off of 10 Å was used and electrostatic interactions were calculated using the partial charges from the FF. An NPT (constant moles, pressure, and temperature) ensemble was used with the Hoover barostat and thermostat;[25] both of which had a time constant of 0.5 ps. Standard conditions of 1 atm and 298 K were used. The systems were initially equilibrated with temperature scaling every 10 steps for 50 ps, so that steady state was achieved. A production run of 10 ns was employed, with a time step of 0.5 fs. The cages were treated in two ways:

- **RBM MD simulations:** All the cages were then treated as rigid bodies, so that the movement of the atoms within each cage was restricted. This was achieved by using the QSHAKE[26] algorithm within DL_POLY 2.20.[26,27] The cages though were permitted to rotate and translate as whole entities and this was controlled by the intermolecular interactions being described using CSFF;
- **FBM MD simulations:** The cage atoms were described explicitly using CSFF, so that their flexibility could be included.

Unless otherwise stated, these were the simulation conditions used throughout this chapter.

4.3. Simulations on the CC1α polymorph

The first polymorph of **CC1** to be characterised experimentally was **CC1α**, and this was recrystallised in the presence of ethyl acetate. Upon desolvation, this occupied the $P2_1/c$ space group. The experimental isotherms for **CC1α** are seen in Figure 1.6a, and these show that it was non-porous to both hydrogen and nitrogen.[22] Simulating the accessible surface area for these gases showed that, although each cage unit possessed an internal cavity, these were disconnected from one another, therefore suggesting that the static structure of **CC1α** is non-porous. The internal cavity of **CC1** can be described as a tetrahedron, and a schematic representation of this is shown in Figure 4.4. As Z'=1 for both **CC1α** and **CC1β**, this tetrahedron is the same for all the cages within each crystal structure. The packing of the cages has a very small influence on the size of the

tetrahedron, with **CC1α** having a larger volume (38.7 Å^3) when compared to **CC1β** (37.8 Å^3). This is because the discrete **CC1** subunits within **CC1α** are compressed very slightly along one axis, distorting the shape of the cage.

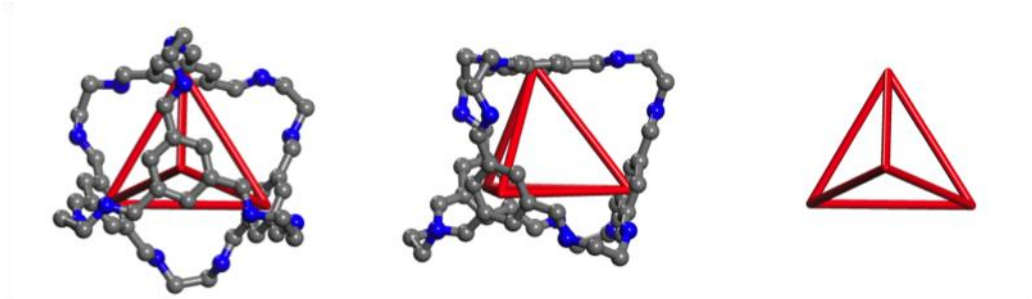


Figure 4.4 – Figure illustrating the **CC1** cage cavity as a tetrahedron (red). Two different angles are shown, as well as the isolated tetrahedron. The tetrahedron is measured by connecting the centre of the four aromatic rings of a cage together.

Each cage cavity has four access points, and the sizes of these are dictated by the diameter of the four cage windows; the diameter of each window was calculated by creating a circle between the three aromatic carbons located on each window. By comparing the size of the window diameter to the diameter of both hydrogen and nitrogen, it is possible to see whether either gas is able to diffuse through these access points. Since each cage has four windows, it is important to consider each one individually.

Figure 4.5 compares these calculated diameters and it is clear that if the crystal packing of the cages is ignored, both polymorphs should allow hydrogen and nitrogen to travel into the cage cavities, because the diameters of the gases (hydrogen = 2.18 Å, nitrogen = 3.10 Å) are smaller than the window diameters for both **CC1α** (3.57-3.95 Å) and **CC1β** (3.62-3.71 Å) windows. This would suggest that any restriction in the diffusion of all the gases through **CC1** is dictated by the crystal packing. The windows in **CC1α** are of unequal sizes, while three of the windows in **CC1β** are equal; this suggests that the crystal packing has an influence on the spherical nature of the cage subunits. The packing of **CC1β** allows the cage to adopt a more spherical shape, while in **CC1α** the cage seems squashed. This indicates that the packing of the cages needs to be taken into consideration.

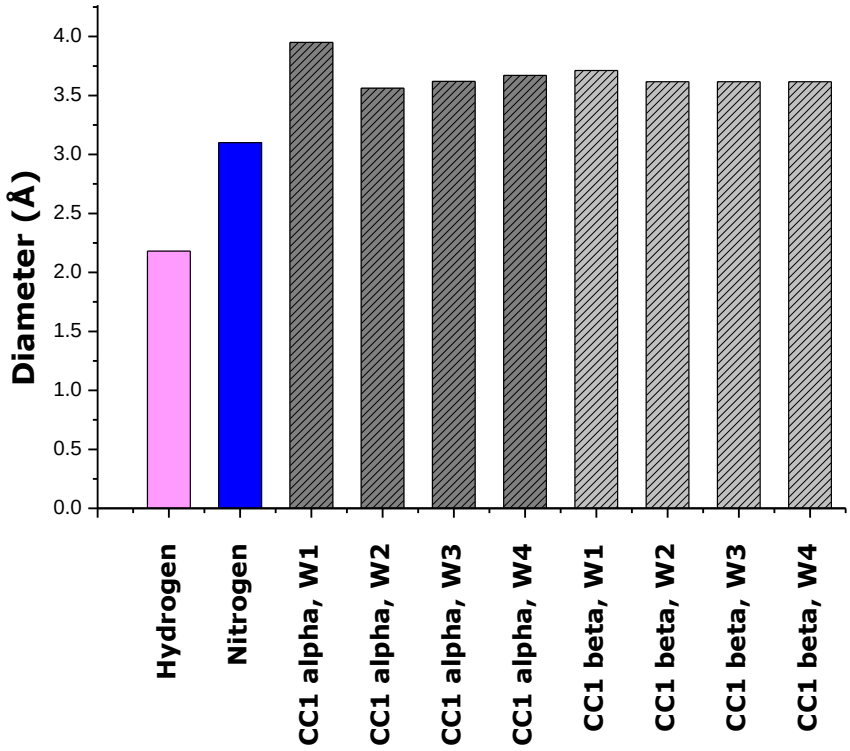


Figure 4.5 – Graph comparing the window diameter of **CC1α** and **CC1β** to the probe diameter of hydrogen and nitrogen. The four windows of each cage are labelled: W1, W2, W3 and W4. The minimum connectivity diameters for the gases have been explained previously in Chapter 2. This shows that both gases are smaller than the access points of an isolated cage therefore have the ability to diffuse into the cage cavity.

In **CC1α** (Figure 4.6), each discrete cage has four 'cage neighbours'. A cage neighbour is the nearest adjacent cage to the particular cage window in question. These were calculated by measuring the close contact distance between the centroid of each cage window, and the centre of the vertices of the neighbouring cage; as defined in Chapter 3. There are two types of cage neighbours:

- **Open cage neighbour**: The cage windows of the two cages are at least partially aligned, thus allowing a direct pore connection between the two cage cavities;
- **Closed cage neighbour**: The cage windows are not aligned, thus prohibiting a direct pore connection between the two cages. The cages could be stacked window-to-arene, or in a window-to-vertex configuration.

The cage windows act as the access points to the cage cavity; hence, the closer the closed cage neighbours are, the more likely they block the access points to the cage cavity. For **CC1α**, it is clear that neighbouring cages block three cage windows, which are all closed cage neighbours, with close contact distances of 2.7 Å, 2.0 Å, and 0.9 Å (Figure 4.6a). In each case, a cage vertex from a

neighbouring cage blocks the cage window. The close proximity of these cages would suggest that a gas molecule would not be able to diffuse directly between these cages; only an extrinsic channel would permit this. However, one window is not as blocked like the others; while the nearest cage is a closed cage neighbour, it is 4.3 Å away from the centre of the cage window (figure 4.6b). This would suggest that there may be potential for gas to permeate between the two cages, as a transient pathway could be formed around the cage vertex. The gas though, would have to be small in size.

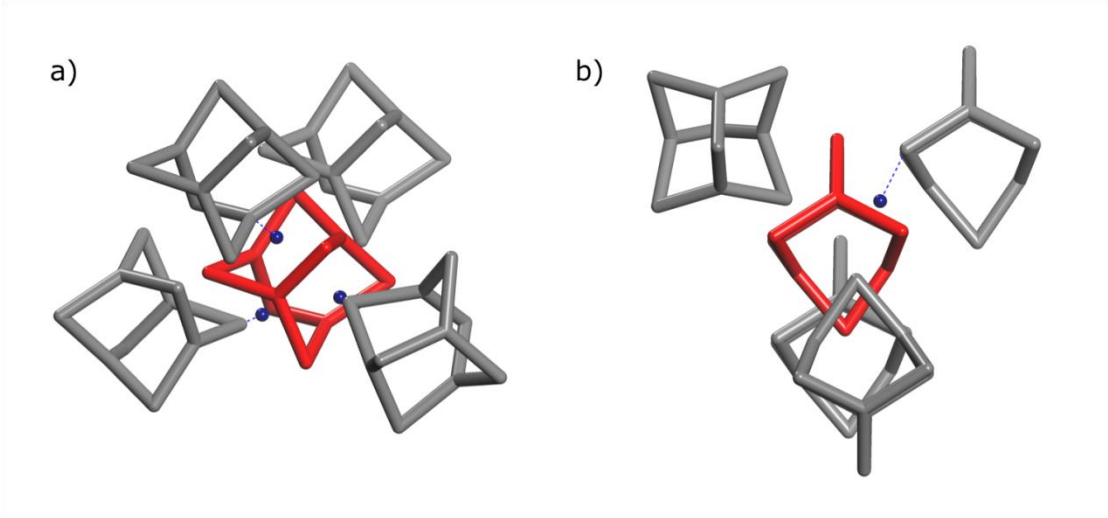


Figure 4.6 – Figure illustrating the packing of **CC1a**. For a given cage (red) there are four cage neighbours (grey). The proximity of these cages are measured from the centroid of the cage windows (blue sphere) and this indicates that for **CC1a** a) there are three cages in close proximity to three cage windows, with b) one cage window more accessible, due to the nearest cage neighbour being 4.3 Å away.

To probe the effect of this crystal packing on the interconnectivity of the voids, the simulated surface area was generated for hydrogen and nitrogen using Zeo^{++} and the guest-free single crystal structures. It was possible to quantify the values for the surface area; these could then be compared to what had been measured experimentally. The simulated surface area is split into two categories; the accessible surface area (ASA), and the non-accessible surface area (NASA). The ASA is the surface area that is fully connected throughout the unit cell, while the NASA is the sum of the isolated regions. If the ASA is calculated to be zero, the system is being classed as non-porous in the classical sense.

Figure 4.7 showed that only NASA is observed for hydrogen and nitrogen, suggesting that it is non-porous to both gases, which agrees with the minimal experimental uptakes for **CC1a**. The surface area in each cage is much smaller for nitrogen. This is unsurprising, because the probe radius used for nitrogen is larger than hydrogen, consequently reducing the space available. Since both

gases are smaller than the diameter of the cage windows, they are able to occupy any space outside the cage cavity that is not blocked by cage neighbours. This explains the elongated shape of the cavities compared to what might be expected for tetrahedral voids centred in the cage molecules.

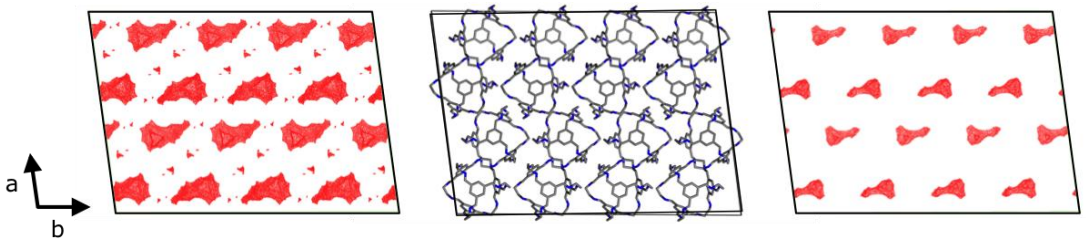


Figure 4.7 – Figure showing the simulated surface area of **CC1a** (centre) for hydrogen (left) and nitrogen (right); both showing only disconnected voids (red). The probe sizes used for hydrogen and nitrogen are 1.09 Å and 1.55 Å respectively.

To simulate the surface area for nitrogen, a number of values have been used in the literature, ranging between the van der Waals radius of a nitrogen atom to the kinetic diameter of nitrogen gas; 1.55 Å to 1.82 Å. These are discussed in more detail in Chapter 2. To show that the methodology for simulating the surface area was accurate, a control, **CC3**, was used. This cage was chosen because it exhibits uptake of both hydrogen and nitrogen (Figure 1.8a and Figure 1.8b), and therefore isolated voids are not expected. Table 4.1 compares the simulated results with experimental SA_{BET}.

A fully crystalline system of **CC3** was reported to have a SA_{BET} of 409 $m^2\ g^{-1}$.[28] Since this experimental SA_{BET} falls in the range of simulated surface areas generated by Zeo^{++} (289 to 586 $m^2\ g^{-1}$), the qualitative comparison of experiment and simulation was deemed valid. The experimental SA_{BET} for **CC1a** was determined to be 23 $m^2\ g^{-1}$, suggesting that **CC1a** is essentially non-porous to nitrogen. This also agrees with the simulated surface areas; the ASA for nitrogen was zero. However, a range of 173 to 388 $m^2\ g^{-1}$ for the NASA showed that the **CC1** internal cavity does give the molecule the capability of being porous; the packing, though, restricts the accessibility of the potential voids. This hypothesis is confirmed by the ASA for hydrogen, which is also zero, with a NASA of 1040 $m^2\ g^{-1}$.

Gas	Probe radius (Å)	CC1α			
		ASA (m^2 g^{-1})	NASA (m^2 g^{-1})	Total (m^2 g^{-1})	Measured SA_{BET} (m^2 g^{-1})
Hydrogen	1.09	0	1040	1040	N/A
Nitrogen[a]	1.55	0	388	388	23
Nitrogen[b]	1.82	0	173	173	23
Gas	**Probe radius (Å)**	**CC1β**			
		ASA (m^2 g^{-1})	NASA (m^2 g^{-1})	Total (m^2 g^{-1})	SA_{BET} (m^2 g^{-1})
Hydrogen	1.09	1665	92	1757	N/A
Nitrogen[a]	1.55	0	737	737	30
Nitrogen[b]	1.82	0	284	284	30
Gas	**Probe radius (Å)**	**CC3**			
		ASA (m^2 g^{-1})	NASA (m^2 g^{-1})	Total (m^2 g^{-1})	SA_{BET} (m^2 g^{-1})
Hydrogen	1.09	1151	92	1244	N/A
Nitrogen[a]	1.55	586	0	586	409
Nitrogen[b]	1.82	289	0	289	409

Table 4.1 – Tabulation of simulated surface areas generated using Zeo^{++} (ASA and NASA) and SA_{BET} taken from previous work.[22,29] [a]the van der Waals radius for nitrogen;[b] the kinetic diameter of nitrogen.

The next step was to simulate the uptake of both hydrogen and nitrogen in **CC1α**. This was achieved using the Sorption tool in Materials Studio to generate GCMC simulated adsorption isotherms for both gases, for comparison to the experimental data. What is interesting is that these disagree, with the simulated uptakes showing an uptake of both hydrogen and nitrogen; six hydrogens per cage (Figure 4.8a), and four nitrogens per cage (Figure 4.8c). This is because the GCMC calculations do not take connectivity into account, and the uptake can be rationalised by adsorption into the internal cavity. This is confirmed by looking at the density maps of the gases sorption sites, Figure 4.8b and Figure 4.8d. It is clear that the majority of uptake is within the cage cavities, and that this is much more defined for nitrogen; this corresponds well to the simulated surface areas in Figure 4.7. For hydrogen, there is a small amount of extrinsic loading and this could suggest that there is space for the gas outside the cage cavity; this was not observed in the nitrogen loading profile (Figure 4.8d). It is important to see whether this extrinsic loading has any impact once the flexibility of **CC1α** is taken into account, as all simulations so far have assumed that the cage molecules are fixed.

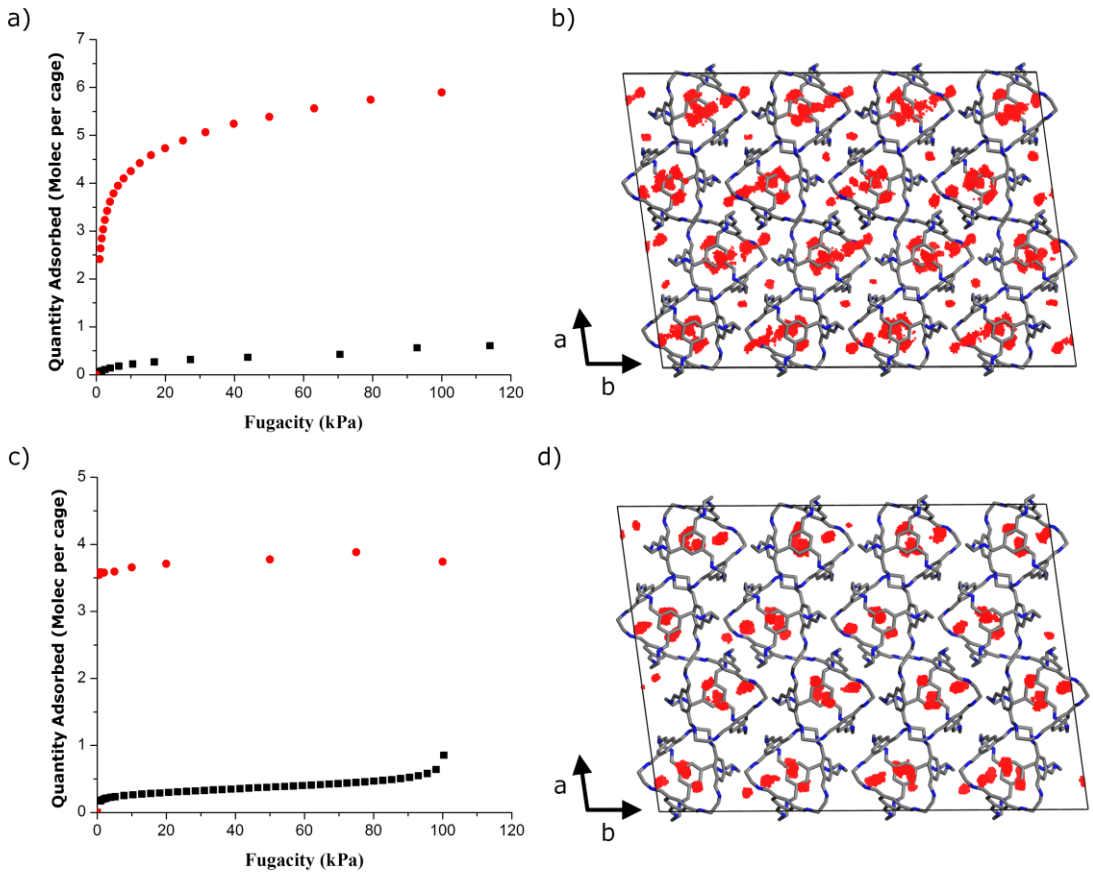


Figure 4.8 - Experimental adsorption isotherms (black) versus the simulated adsorption isotherms (red) for a) hydrogen in **CC1a** and, c) nitrogen in **CC1a**. The density maps of the sampled gas loading positions during the GCMC calculations are shown for, b) hydrogen and, d) nitrogen. The hydrogen atoms on the cages have been omitted for clarity.[22] This was simulated at 77 K.

To quantify the importance of framework flexibility for **CC1a**, both RBM and FBM MD simulations were carried out. Now that framework flexibility has been taken into consideration, it was possible to see whether, in contradiction to the experimental results, gases diffused through the system. Figure 4.9 shows the superimposition of the gas trajectory for each MD simulation. The RBM MD simulations for hydrogen and nitrogen in **CC1a** (Figure 4.9a and Figure 4.9c) show that although the gases are able to move through the cage windows, no channels are formed between the cages during the simulation, indicating that the packing of **CC1a** is restricting the movement of the gases. The FBM MD simulation for nitrogen is also restricted to a single, starting cage. This agrees with all simulated and experimental data thus far, confirming that **CC1a** is non-porous to both hydrogen and nitrogen.

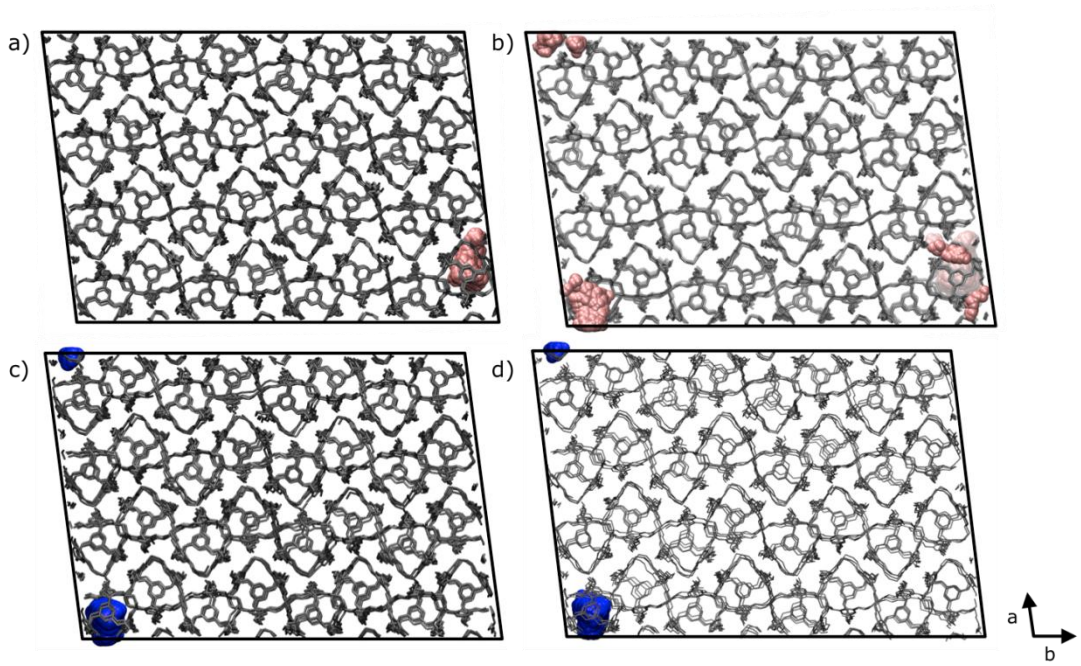


Figure 4.9 – A superimposition of the gas molecules' positions during MD simulations on **CC1α**; a) RBM MD simulations and b) FBM MD simulations for hydrogen. c) The RBM MD simulations and d) FBM MD simulation for nitrogen. All the cages are coloured grey, with hydrogens omitted for clarity. Hydrogen gas is coloured pink and nitrogen gas blue. The total simulation length was 10 ns, with a frame rate of 1 ps and a temperature of 298 K.

The results for the FBM MD simulation for hydrogen showed something different; the gas occupied two cages during the simulation (Figure 4.10b). The gas hopped once during the 10 ns simulation, from one cage to another, and this was facilitated by one cage neighbour being further away than the other three, thus creating a small extrinsic void adjacent to one window. This has been explained earlier, and can be seen schematically in Figure 4.4. This acts as a connection point between the two cages so that, even though they are closed cage neighbours, a pathway could be formed. The extra space around the cage window allows for gas diffusion between these two cages; this diffusion out of a cage window, though, is a rare event (occurring only once over the 10 ns simulation). The close proximity of the other closed cage neighbours prevents the gas from diffusing further through the rest of the system, therefore rendering **CC1α** non-porous for hydrogen when considered as a bulk, three-dimensional crystal.

Overall, the simulations showed that **CC1α** is non-porous to both hydrogen and nitrogen, and this agrees with experiment. The surface area calculations and simulated GCMC sorption isotherms, however, initially suggested that the discrete **CC1** unit might be porous to both gases. As revealed by MD, though, the cavities in **CC1α** remain isolated as pockets of free volume as a result of the closed cage neighbours in the **CC1α** polymorph prevented any extended connectivity,

restricting gas permeability throughout the system. The next step was to look at the second **CC1** polymorph, **CC1β**, to see whether the selective uptake of hydrogen over nitrogen could be rationalised.

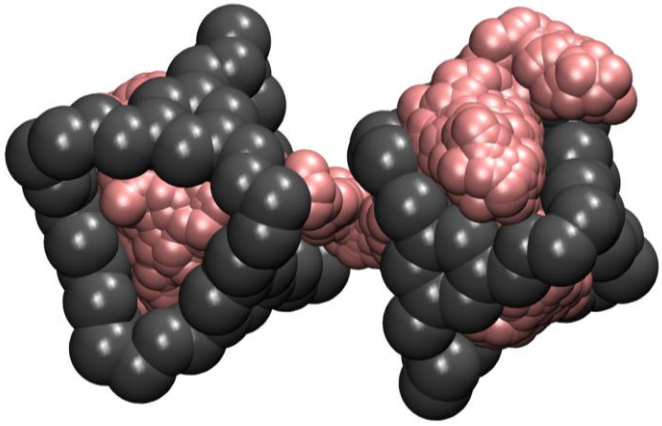


Figure 4.10 – Figure showing the superimposition of hydrogen during the FBM MD simulations for **CC1α**. This shows that there is diffusion from one cage to another, though this is limited between two cages. Hydrogen gas is coloured pink and the cages grey. The hydrogens on the cage have been omitted for clarity.

4.4. Simulations on the CC1β polymorph

The **CC1β** polymorph adopts the *R*3 space group after being recrystallised in the presence of dichloromethane and subsequently desolvated. The experimental uptakes, Figure 1.6b, show that this polymorph is selectively porous to hydrogen over nitrogen. To rationalise this, first the simulated surface areas were generated using Zeo^{++}. Figure 4.11 shows that when a probe radius of 1.09 Å is used to produce the surface area for hydrogen, a fully connected pore topology is generated. This would suggest that hydrogen would be able to diffuse through the system freely, and therefore uptake of hydrogen gas would be expected. When the nitrogen surface area was simulated using a probe of 1.55 Å, the internal cavities in **CC1** were observed, as well as additional void space pointing out of three of the four cage windows, but these voids were again disconnected. This would suggest that the packing of the cages was restricting the connectivity.

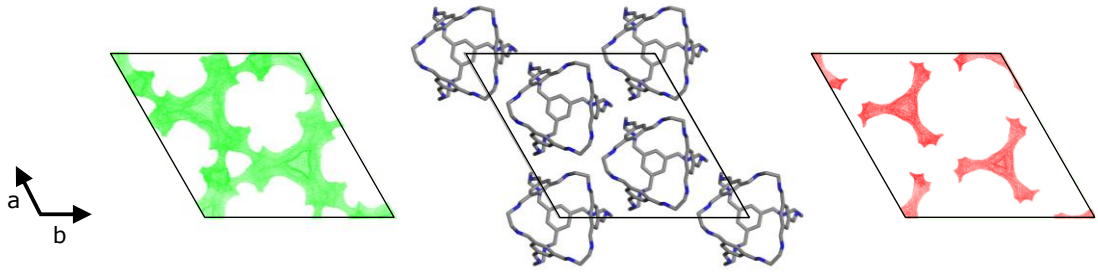


Figure 4.11 – Figure showing the simulated surface area of **CC1β** (centre) for hydrogen (left) and nitrogen (right). This shows that **CC1β** has a connected network for hydrogen (green), but is still disconnected to nitrogen (red). The probe sizes used for hydrogen and nitrogen are 1.09 and 1.55 Å respectively.

The packing motif of **CC1** in the **CC1β** polymorph was therefore examined (Figure 4.12). This showed that the crystal packing of the **CC1β** polymorph differed greatly to the **CC1α** packing motif. First, three of the cage windows, the access points to the cage cavity, were identical, each having two open cage neighbours that were 4.8 Å away (Figure 4.12b). These open contact neighbours are much further apart than any of the cage neighbours observed in the **CC1α** polymorph, and this would suggest that direct diffusion between them might be possible. In addition, the nearest closed cage neighbour for the final window is stacked directly above it, in a window-to-arene configuration, with a much shorter distance of only 2.5 Å. This is similar to **CC1α**, and would suggest that diffusion to this cage is not possible, at least *via* a pathway between the cages. The packing of **CC1** in the **CC1β** motif would imply a two-dimensional pore structure where each layer could only be connected *via* an extrinsic channel.

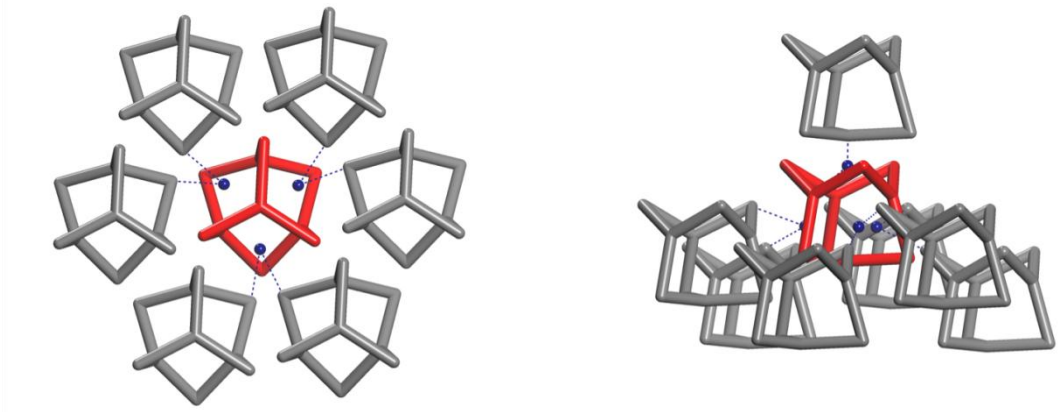


Figure 4.12 - Figure illustrating the packing of **CC1β**. For a given cage (red) there are seven cage neighbours (grey)**;** a) for three of the windows the cage neighbours are identical; two open cage neighbours 4.8 Å apart. b) The other cage window has a closed cage neighbour stacked directly above it and is only 2.5 Å away, suggesting that a direct diffusion pathway is blocked.

The packing mode of **CC1β** helps to rationalise the selective uptake of hydrogen over nitrogen. It was also possible to compare the simulated ASA and NASA for nitrogen in **CC1β**, and these are given in Table 4.1. The ASA, like **CC1α**, was zero, with a NASA ranging from 284 to 737 $m^2\ g^{-1}$; this matches well with the experimentally measured SA_{BET} of 30 $m^2\ g^{-1}$. The ASA for hydrogen was calculated to be 1665 $m^2\ g^{-1}$, with a NASA of just 92 $m^2\ g^{-1}$ showing that a decrease in size has a large effect on the connectivity of the **CC1β** polymorph.

The Sorption package within Materials Studio was used to generate GCMC simulated adsorption isotherms for both gases within **CC1β**, and the results were then compared to the experimental data. The simulated adsorption isotherm for hydrogen showed uptake, but this was an overprediction compared to the experimental observation (Figure 4.13a). This may have been caused by a number of factors. First, CSFF was used as the FF during the GCMC calculations. Although all the internal degrees of freedom had been parameterised for these imine cage structures, the interatomic interactions between the discrete cage units had been scaled appropriately, and the gases had been modelled using literature parameters without the specific gas-cage interactions being reparameterised, as this was beyond the scope of the FF development. As a result, these are fitted using the generic interactions found within PCFF, as this is the FF CSFF was adapted from. Another reason for the overprediction of gas uptake might be the lack of crystallinity in the sample. It has been shown recently that the crystallinity of the experimental sample has a large influence on gas uptake;[29] this is because the contribution of extrinsic porosity can sometimes be increased if the crystallinity of the sample is lower. This can either increase or decrease the surface area, and is dependent on the connectivity of the voids. If **CC1β** possessed any amorphous character this could make the comparison with theory inaccurate. Although the simulated uptake is overpredicted, the fact that **CC1β** exhibits hydrogen uptake would suggest that pore connectivity within the crystal structure is observed.

The loading profile for hydrogen, Figure 4.13b, differed slightly to the surface area plots (Figure 4.11), in so much as the gas is loaded into discrete pockets opposed to a fully connected pore network. These pockets include and extend out of the internal cavity found within each cage, through the cage window, and are only limited by the crystal packing. This suggests that any connections between the cages are too small for gas molecules to be loaded into, and act solely as pathways for hydrogen to 'hop' through. MD simulations should show that, once flexibility is included, diffusion of hydrogen is observed through these

connections. The simulated sorption for nitrogen contradicts the experimental results, suggesting that nitrogen uptake is possible. This can be rationalised in the same manner as for **CC1α**; the intrinsic cage cavities possess the capability to store nitrogen, explaining the simulated uptake, but the cavities are disconnected therefore making diffusion into these voids impossible.

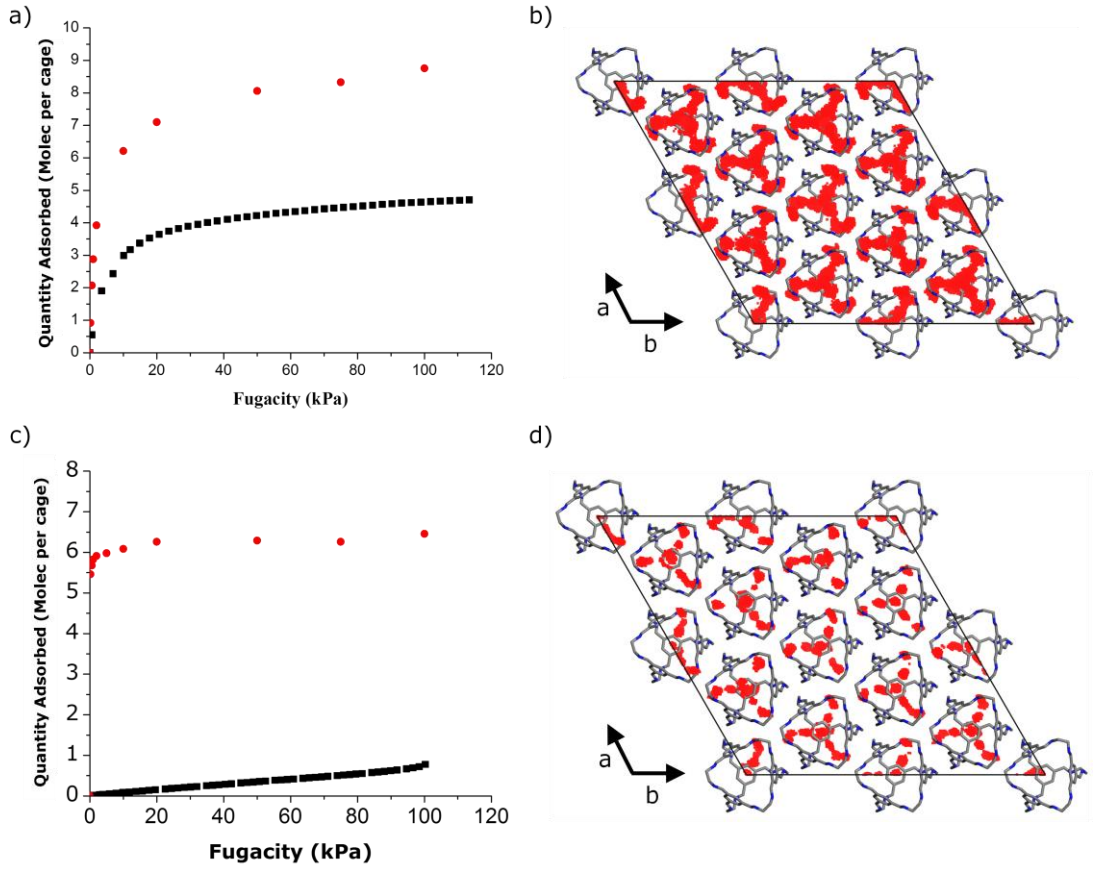


Figure 4.13 - Experimental adsorption isotherms (black) versus simulated adsorption isotherms (red) for a) hydrogen in **CC1β** and c) nitrogen in **CC1β**. The density maps of the sampled loading during the GCMC calculations are shown for b) hydrogen and d) nitrogen. The hydrogen atoms on the cages have been omitted for clarity.[22] This was simulated at 77 K.

By introducing flexibility, the permeability of gases through **CC1β** can be examined; again both RBM and FBM MD simulations were performed. The results showed diffusion of hydrogen through **CC1β**, while nitrogen was restricted to a single cage. Figure 4.14 shows the superimposition of the gas trajectory for each simulation. These results tie in well with the observed experimental uptakes, as well as the simulated surface areas. The results from the hydrogen MD simulations indicate that by increasing the flexibility of **CC1β**, more cages are occupied during the simulation. During the RBM MD simulation, only 37% of the cages were occupied, whilst 96% of the cages were occupied in the FBM MD

simulation. This suggests that the diffusion coefficient for the FBM MD simulations would be faster than for the RBM MD simulations.

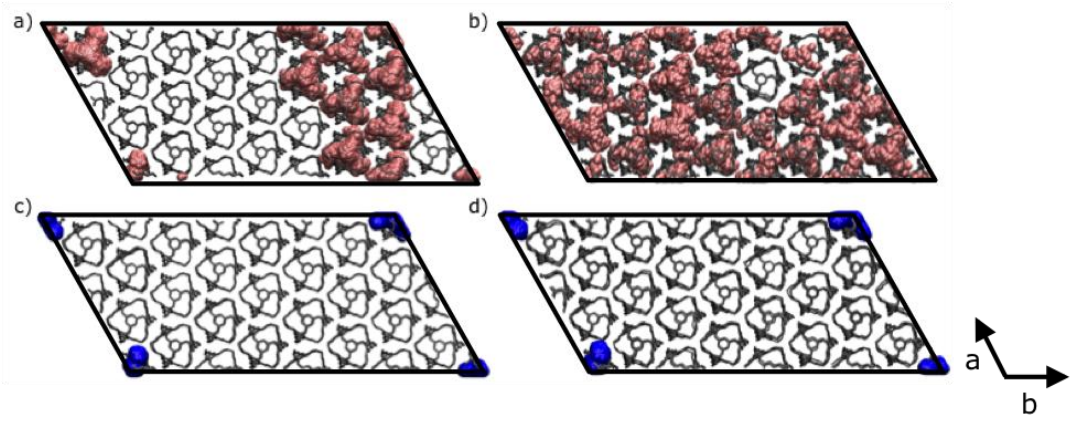


Figure 4.14 – A superimposition of the gas molecules' positions during MD simulations on **CC1β**; a) RBM MD simulations and b) FBM MD simulations for hydrogen. c) The RBM MD simulations and d) FBM MD simulation for nitrogen. All the cages are coloured grey, with hydrogens omitted for clarity. Hydrogen gas is coloured pink and nitrogen gas blue. The total simulation length was 10 ns, with a frame rate of 1 ps and a temperature of 298 K.

The mean squared displacement of hydrogen for the RBM and FBM MD simulations was plotted as a function of time, Figure 4.15. The slope was used to calculate the self-diffusivities of hydrogen. For normal diffusion to occur the slope of the logarithmic plot must be close to 1; this is the case for both simulations. The diffusion coefficient of hydrogen in **CC1β** was calculated to be much faster for the FBM MD simulations at 7.82 x 10^{-9} m^2 s^{-1}, when compared to 4.14 x 10^{-10} m^2 s^{-1} for the RBM MD simulation. Thus, by including the flexibility of the cages in the simulation, the diffusion coefficient was nearly twenty times faster. It is important to note that the increased flexibility of the FBM MD simulation was not necessary to observe hydrogen diffusion in **CC1β**, as this was seen in the RBM MD simulations, suggesting that no breathing motion of the cage windows is necessary for hydrogen to pass through them.

An example of the diffusion pathway of hydrogen between two cages in **CC1β** is shown in Figure 4.16. There is a distinct channel between the two cages, but with no interstitial site in which the gas resides for longer periods. In addition, the gas attempts briefly to diffuse in another direction, only to move back into the original cage, as its pathway was blocked. When compared to the isolated pockets seen in density plots (Figure 4.16b), and the narrow connections seen in the surface area plots for hydrogen (Figure 4.11), this would confirm that although narrow channels do exist between the cages, these are not large enough for hydrogen to remain for long periods of time. This is confirmed when the density maps for **CC1β** are examined; there are no observed sorption sites for

these channels. When the packing motif is considered (Figure 4.12), this implies that direct cage-to-cage diffusion is only seen along one plane. The FBM MD simulations showed that hydrogen occupied 96% of the cages during the 10 ns simulation, and hence two-dimensional planes must become at least transiently connected. Since one cage window is blocked by another cage stacked directly above it, any three-dimensional connectivity must be facilitated *via* an extrinsic pathway. This window-to-arene stacking configuration is shown in Figure 4.12.

Overall, the simulations for **CC1β** rationalise nicely the selective diffusion of hydrogen over nitrogen, by showing a connected pore topology is observed only for hydrogen. The MD simulations also show that hydrogen is able to diffuse through the system, whilst nitrogen is confined to a single cage. This again demonstrates how important the packing of the **CC1** subunits is in determining the interconnectivity of voids and hence porosity.

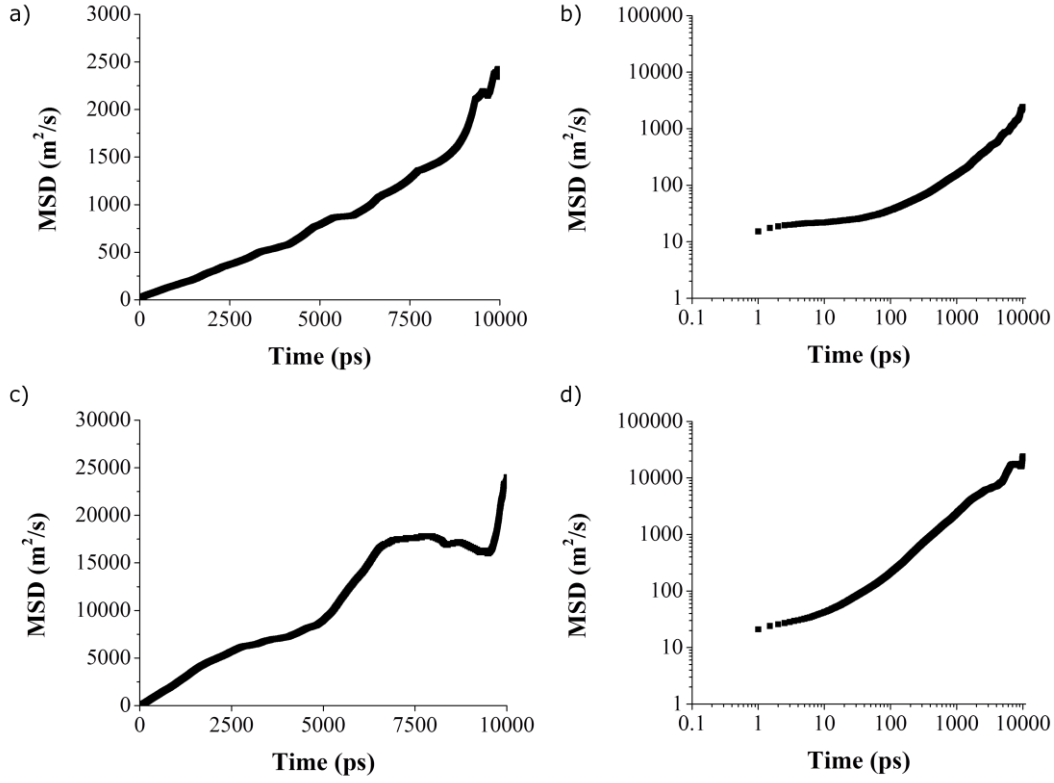


Figure 4.15 – The MSD for hydrogen in **CC1β** for a) the RBM simulation, b) the logarithmic plot that confirms normal diffusion has occurred, c) the FBM MD simulation, and d) logarithmic plot for the FBM MD simulation.

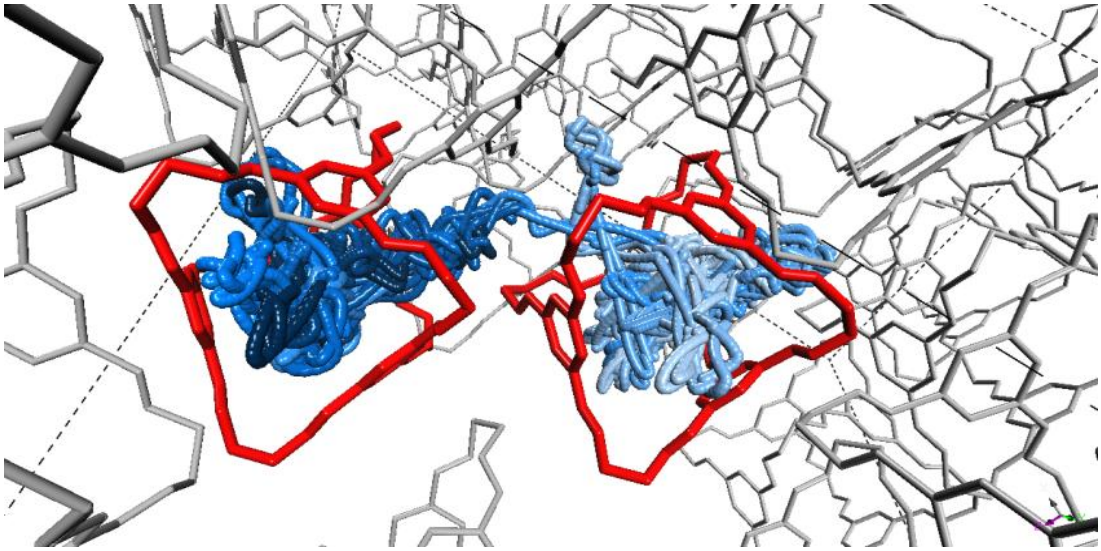


Figure 4.16 – Superimposition of hydrogen during the FBM MD simulations for **CC1β**. This shows the diffusion pathway for hydrogen between two cages (highlighted in red). The hydrogen has been coloured blue and the trajectory has been coloured with respect to time – the colour darkening as time progresses. Hydrogen atoms on the cages have been omitted for clarity.

4.5. Conclusion

The simulations carried out in this chapter have helped to rationalise known experimental results, as well as to provide further insight into how and why polymorphism in the structure of **CC1** alters the porosity of these molecular crystals. Each molecule of **CC1** has its own internal cavity that can be represented as a tetrahedron. Both hydrogen and nitrogen are small enough to diffuse into this cavity *via* the four access points, the cage windows, and the connectivity of these voids is dictated by the packing of the discrete units. In **CC1α**, the discrete units are packed such that the voids remain isolated from one another. MD simulations rationalised that although sorption into these cavities was observed in the GCMC simulations, there was no way for the gas to diffuse from cage to cage; thus minimal uptake is seen experimentally. In the **CC1β** polymorph, discrete units of **CC1**, crystallise such that it is selectively porous to hydrogen over nitrogen. This selectivity can be explained by consideration of the pore structure and connectivity; a small decrease in gas size (from a radius of 1.55 to 1.09 Å) is enough to switch the pores from being non-porous to porous, hence rationalising the selective porosity of **CC1β**. This subtle change is validated by the MD simulations, which showed that hydrogen is free to diffuse throughout the whole system, while nitrogen is confined to a single cage. Using this information, schematic pore structures for **CC1α** and **CC1β** with respect to hydrogen and nitrogen can be constructed, as shown in Figure 4.17 and Figure

4.18 respectively. The pore structure for **CC1α** shows that for both hydrogen and nitrogen, no connectivity is observed.

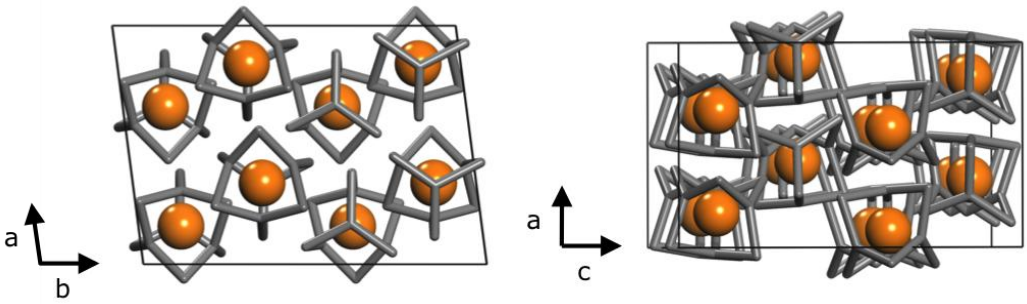


Figure 4.17 – Computationally determined pore connectivity for hydrogen and nitrogen in **CC1α**: only isolated voids (orange) are present in **CC1α** and this restricts the diffusion of both hydrogen and nitrogen through the system resulting in **CC1α** being non-porous to both gases.

CC1β is also non-porous to nitrogen, as only isolated voids are present (Figure 4.18b). However, there is potential for hydrogen uptake in **CC1β** and this is facilitated by a fully interconnected three-dimensional pore structure (Figure 4.18a). The packing motif of **CC1β** (Figure 4.12) suggests two-dimensional planes, with an extrinsic channel linking them, and the MD simulations further validated this. This extrinsic pathway is shown in Figure 4.18a. Although the structure of **CC1γ** was not examined in this work, the experimental uptake of both hydrogen and nitrogen (Figure 4.2) would suggest that the pore topology of this system should be fully connected for both gases.

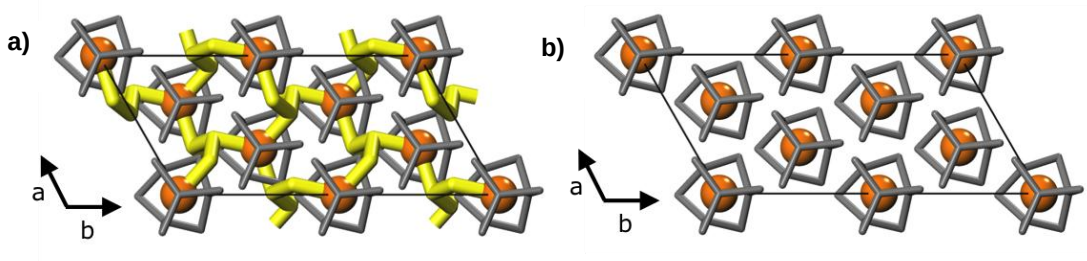


Figure 4.18 – Computationally determined pore connectivity for hydrogen and nitrogen in **CC1β**: a) the isolated voids in **CC1β** are connected for hydrogen, with the layers being connected *via* an extrinsic pore pathway, while b) shows that only isolated voids (orange) are present in the pore structure of **CC1β** for nitrogen; this restricts the diffusion of through the system. This results in **CC1β** being porous to hydrogen and non-porous to nitrogen.

GCMC calculations consistently overestimated the gas uptake of the polymorphs, and this can be explained by a number of factors. First, when crystallised in their non-porous state, discrete void space is available and the simulations load gas into these cavities, thus generating results that contradict experimental observations, as there is no consideration of a connectivity pathway for them. If this work were to be repeated, it would be beneficial to add an additional criterion

to the GCMC simulations that only permitted gas sorption into areas found on the interconnected pore structure. This would automatically rule out any sorption in **CC1α** and nitrogen sorption in **CC1β**, thus reproducing experimental findings. In addition, the simulated uptake of hydrogen in **CC1β** was overestimated; this could be because of the quality and crystallinity of the sample used, as this has been shown to have a major affect on the uptakes of other cages.[28] More likely is that the FF used, CSFF, does not correctly describe the intermolecular interactions between the cages and gases, which were not explicitly parameterised. This may have played a role in the overestimation of the results. Overprediction of gas uptake in GCMC simulations is a common problem for porous materials and has been well documented in the literature for MOFs.[30-34] If this work was extended, it would be prudent to tailor the interactions between the gases and the cages. This could also provide additional information about adsorption energies and potential binding sites.

An advantage of porous cages is their ability to crystallise into different polymorphs. This opens up the potential for crystal structure prediction (CSP), and work has already been published that post-rationalises the polymorphs of other cages.[11] It would be conceivable to use CSP to generate a whole subset of hypothetical low energy polymorphs for **CC1**, and subsequently to use the knowledge gained from this work to predict their properties. Pore connectivity, sorption, diffusion, and surface areas could be used as screening criteria, and subsequently candidates chosen for targeted synthesis. The ability to recrystallise **CC1** with many different solvents also opens up the potential for high throughput studies.

The idea of screening porous cages for specific properties is focused on in detail in the next chapter, which deals with the cage **CC3** and a variety of different gases.

4.6. References

(1) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. Nature. **2003**, *423*, 705–714.

(2) Atwood, J. L.; Barbour, L. J.; Jerga, A. Angew. Chem. Int. Ed. Engl. **2004**, *43*, 2948–2950.

(3) Kuznicki, S. M.; Bell, V. A.; Nair, S.; Hillhouse, H. W. Nature. **2001**, *412*, 720-724.

(4) Walz, T.; Hirai, T.; Murata, K.; Heymann, J. B.; Mitsuoka, K. Nature. **1997**, *387*, 624-627.

(5) Kitagawa, S.; Kitaura, R.; Noro, S. Angew. Chem. Int. Ed. **2004**, *43*, 2334–2375.

(6) MacKinnon, R. Angew. Chem. Int. Ed. **2004**, *43*, 4265 – 4277.

(7) Jiang, S.; Bacsa, J.; Wu, X.; Jones, J. T. A.; Dawson, R.; Trewin, A.; Adams, D. J.; Cooper, A. I. Chem. Commun. **2011**, *47*, 8919.

(8) Jones, J. T. A.; Holden, D.; Mitra, T.; Hasell, T.; Adams, D. J.; Jelfs, K. E.; Trewin, A.; Willock, D. J.; Day, G. M.; Bacsa, J.; Steiner, A.; Cooper, A. I. Angew. Chem. Int. Ed. Engl. **2011**, *50*, 749–753.

(9) Wadell, H. J. of Geo. **1935**, *43*, 250-280.

(10) Merkel, T. C.; Freeman, B. D.; Spontak, R. J.; He, Z.; Pinnau, I. Science **2002**, *296*, 519-522.

(11) Jones, J. T. A.; Hasell, T.; Wu, X.; Bacsa, J.; Jelfs, K. E.; Schmidtmann, M.; Chong, S. Y.; Adams, D. J.; Trewin, A.; Schiffman, F.; Cora, F.; Ben Slater; Steiner, A.; Day, G. M.; Cooper, A. I. Nature. **2011**, *474*, 367–371.

(12) Bradshaw, D.; Warren, J. E.; Rosseinsky, M. J. Science. **2007**, *315*, 977–980.

(13) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Microporous Mesoporous Mater. **2012**, *149*, 134–141.

(14) Serre, C.; Mellot-Draznieks, C.; Surblé, S.; Audebrand, N.; Filinchuk, Y.; Férey, G. Science. **2007**, *315*, 1828–1831.

(15) Horike, S.; Shimomura, S.; Kitagawa, S. Nat. Chem. **2009**, *1*, 695–704.

(16) Accelrys; Accelrys Software, Inc.: San Diego, **2011**.

(17) Rabone, J.; Yue, Y. F.; Chong, S. Y.; Stylianou, K. C.; Bacsa, J.; Bradshaw, D.; Darling, G. R.; Berry, N. G.; Khimyak, Y. Z.; Ganin, A. Y.; Wiper, P.; Claridge, J. B.; Rosseinsky, M. J. Science. **2010**, *329*, 1053–1057.

(18) Serre, C.; Bourrelly, S.; Vimont, A.; Ramsahye, N. A.; Maurin, G.; Llewellyn, P. L.; Daturi, M.; Filinchuk, Y.; Leynaud, O.; Barnes, P.; Férey, G. Adv. Mater. **2007**, *19*, 2246–2251.

(19) Atwood, J. L. Science. **2002**, *296*, 2367–2369.

(20) Atwood, J. L. Science. **2002**, *298*, 1000–1002.

(21) Flippen, J. L.; Karle, J.; Karle, I. L. J. Am. Chem. Soc. **1970**, *92*, 3749–3755.

(22) Jones, J. T. A.; Holden, D.; Mitra, T.; Hasell, T.; Adams, D. J.; Jelfs, K. E.; Trewin, A.; Willock, D. J.; Day, G. M.; Bacsa, J.; Steiner, A.; Cooper, A. I. Angew. Chem. Int. Ed. Engl. **2011**, *50*, 749–753.

(23) Barbour, L. J. Chem. Commun. **2006**, 1163.

(24) Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. Nat. Mat. **2009**, *8*, 973–978.

(25) Melchionna, S.; Ciccotti, G.; Lee Holian, B. Mol. Phys. **1993**, *78*, 533–544.

(26) Smith, W.; Forester, T. R. J. Mol. Graph. **1996**, *14*, 136–141.

(27) Forester, T. R.; Smith, W. Daresbury Laboratory, **2000**.

(28) Hasell, T.; Chong, S. Y.; Schmidtmann, M.; Adams, D. J.; Cooper, A. I. Angew. Chem. Int. Ed. Engl. **2012**, *51*, 7154–7157.

(29) Hasell, T.; Chong, S. Y.; Jelfs, K. E.; Adams, D. J.; Cooper, A. I. J. Am. Chem. Soc. **2012**, *134*, 588–598.

(30) Parkes, M. V.; Staiger, C. L.; Perry, J. J., IV; Allendorf, M. D.; Greathouse, J. A. Phys. Chem. Chem. Phys. **2013**, *15*, 9093.

(31) Thornton, A. W.; Furman, S. A.; Nairn, K. M.; Hill, A. J.; Hill, J. M.; Hill, M. R. Microporous Mesoporous Mater. **2013**, *167*, 188–197.

(32) Dubbeldam, D.; Frost, H.; Walton, K. S.; Snurr, R. Q. Fluid Phase Equilib. **2007**, *261*, 152–161.

(33) Düren, T.; Bae, Y.-S.; Snurr, R. Q. Chem. Soc. Rev. **2009**, *38*, 1237–1247.

(34) Li, W.; Grimme, S.; Krieg, H.; Möllmann, J.; Zhang, J. J. Phys. Chem. C. **2012**, *116*, 8865–8871.

Chapter 5

Modelling the diffusion of gases through CC3 using molecular dynamics

5.1. Introduction

Previously, it has been shown that tetrahedral organic cages can crystallise to form porous molecular crystals that adsorb small guest molecules, such as gases.[1,2] As most organic molecules pack in such a way to minimise free space, exhibit minimal void volume in the solid state was rare. Porous organic cages, by contrast, are permanently porous to a variety of gases. Study of the static structure alone does not adequately explain the porosity in these materials; this was especially evident for **CC3**, which shows gas uptake for a variety of gases. Computational analysis of the static structure alone conflicts with what was observed experimentally. We therefore chose to employ molecular dynamic simulations to understand these solids.

To do this, the diffusion of seven gases, varying in shape and size, were simulated within a test system, **CC3**-*R*; only the heterochiral system was examined, therefore this was referred to as **CC3** throughout. Initially, a rigid body approximation[3-12] was used in combination with molecular dynamics (MD), to see which gases would diffuse through the system; here, the intramolecular potentials were restricted with just the intermolecular potentials being described using CSFF. Subsequently, MD simulations were performed on the same systems, but this time cage subunits were treated as fully flexible.

It was important to see if the experimentally determined properties for **CC3** can be reproduced using molecular simulations. This allows us to make a judgement on whether the simulations are representative; if this was the case, analysis of the simulations would yield insight into the types of diffusion taking place, as well as providing an appreciation of the kinetics involved.

Unlike many other porous network systems, **CC3** is comprised of molecular subunits, and these pack in a way to generate a 3D diamondoid pore structure; these pores have a similar size to a number of small gases, for example krypton and xenon. A detailed description, along with the experimental sorption uptakes, can be found in Chapter 1.

5.2. Void analysis of static CC3 crystal structure

It was possible to examine the different pore topologies in **CC3** with respect to various gases, allowing an insight into the extent that these gases were accessible to the **CC3** pore structure when the cages were 'static'. This can be achieved by altering the probe radii used to generate the surface areas, so that each shows the solvent surface area for the given gas. The accessible surface

area of **CC3** was generated using Zeo^{++} and the static X-ray crystal structure for each gas by altering the probe radius used.[13,14] Figure 5.1 shows the different pore topologies for the seven chosen gases: hydrogen, nitrogen, carbon dioxide, methane, sulfur hexafluoride, krypton, and xenon. For clarity, Figure 5.2 shows how the pore topology fits within a cage molecule within the system.

The pore topologies generated show 3D diamondoid networks for hydrogen, nitrogen, and carbon dioxide, whereas methane, krypton and xenon exhibit disconnected voids. This 'static' view suggests that these gases would not diffuse through the **CC3** system. However, in the case of krypton, the diameter of the gas and the window diameter in the static **CC3** structure were very similar. Hence, this conclusion may be misleading as there were a number of different values for the diameter for krypton that have been used, as discussed earlier (Chapter 3). In any case, the similarity suggested scope for molecular separation of krypton, as described briefly in Section 5.10.

By using the smallest value, reported by Fernandez *et al.*,[15-17] there was a continuous 3D diamondoid network for krypton, Figure 5.3. The difference between the window diameter of **CC3** and the diameter used for krypton was just 0.016 Å, therefore there was potential for diffusion through the system, as the gases were not modelled as hard spheres. Normally this effect would be negligible, but for krypton this may be enough to allow it to diffuse through the system.

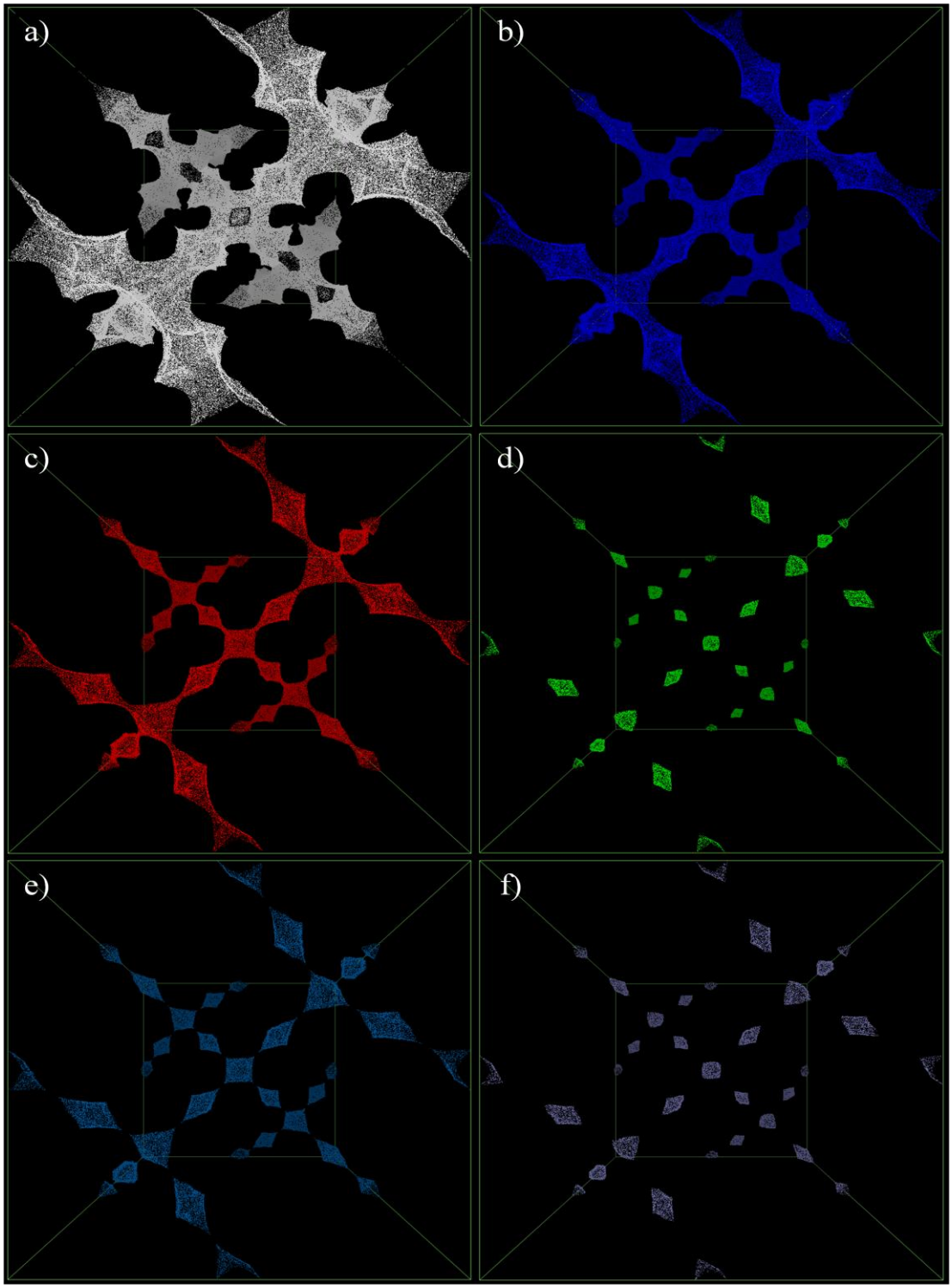


Figure 5.1 – Connectivity of **CC3** pore topology for various gases; a) hydrogen (1.09 Å), b) nitrogen (1.55 Å), c) carbon dioxide (1.70 Å), d) methane* (2.126 Å), e) krypton (1.818 Å) and f) xenon (2.05 Å). Probe radii used are in brackets.

* This assumes that methane behaves as a sphere; details for this in Section 3.10.

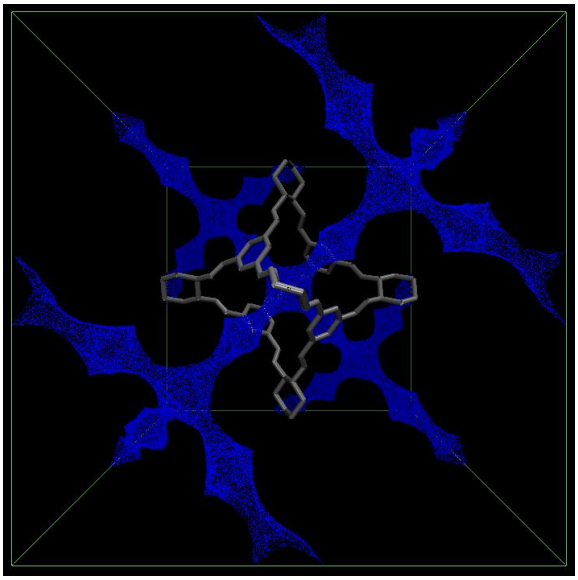


Figure 5.2 – A single cage within the **CC3** crystal structure and an example of a pore structure. For clarity, hydrogens have been removed and the cage coloured grey.

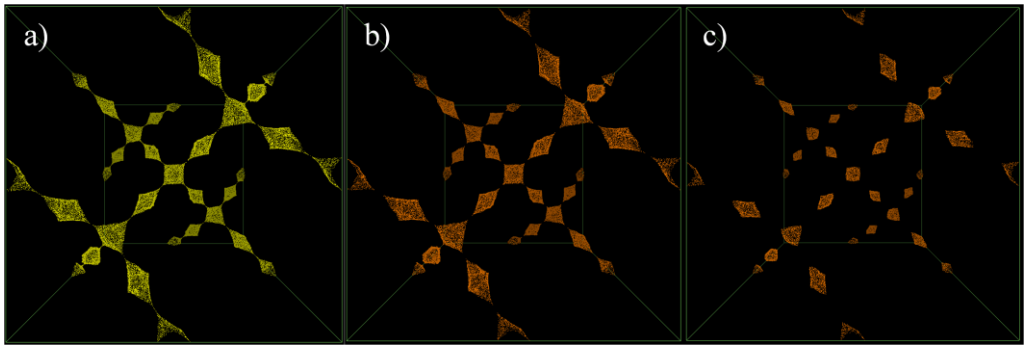


Figure 5.3 – Pore connectivity for krypton using different literature values for the Van der Waal radii. From left to right; 1.818 Å,[3-12,18] 1.8 Å[1,16] and 2.02 Å.[19] If the voids were coloured yellow, the channels were connected, if orange, only disconnected voids were present.

By comparing the pore topologies of the various gases, it was possible to suggest that **CC3** could potentially be a good material for gas separation. Initial findings point to **CC3** allowing the diffusion of hydrogen, nitrogen, carbon dioxide, while preventing the diffusion of methane, xenon and sulfur hexafluoride; diffusion of krypton would be marginal. However, experimental sorption isotherms for krypton show that at a pressure of 1 bar and a temperature of 263 K, there is a sharp uptake of 1.8 mmol g^{-1}. This however would suggest that representing **CC3** as a static structure, when considering its porosity, was an invalid assumption. This would also suggest that work done by Sikora *et al.*[18] on hypothetical MOFs could not be directly transferred to porous organic cages, as the size of the pores was too close to that of the gases. Unlike some larger more rigid MOFs,[20,21] flexibility of both the individual components, as well as their relationship to one another, could be instrumental in their gas sorption properties. The ambition of

modelling various gases in **CC3** was to see if an accurate representation of their diffusion could be attained, the style of diffusion taking place unravelled, and any advantageous characteristics of **CC3** highlighted; for example by molecular screening of various small molecules. Within this research, two approaches were undertaken. The rigid body method (RBM) and the flexible body method (FBM); these were described in detail in the previous chapter (Section 4.2).

5.3. Simulation Details

MD simulations were performed to look at the diffusion properties of seven different gases, hydrogen, nitrogen, carbon dioxide, methane, krypton, xenon, and sulfur hexafluoride. A 2 x 2 x 2 supercell of **CC3**, containing several thousand atoms, was used to represent the crystal structure; this has the initial cell dimensions A = B = C = 49.59 Å, α = β = Y = 90°. In this, a single molecule of the desired gas was inserted such that it was at the centre of one of the cages. A potential cut off of 10 Å was used, and electrostatic interactions were calculated using the partial charges from the FF. NPT (constant moles, pressure, and temperature) ensemble dynamics with the Hoover barostat[22] and thermostat[22] were used throughout, and both had a time constant of 0.5 ps. Standard conditions of 1 atm and 298 K were used. The systems were initially equilibrated with temperature scaling every 10 timesteps for 50 ps, so that steady state was achieved. A production run of 10 ns was run employing a time step of 0.5 fs. The cages were treated in two ways:

1. All the cages were then treated as rigid bodies, so that the movement of the atoms within each cage was restricted. This was achieved by using the QSHAKE algorithm within DL_POLY 2.20.[23] The cages though were permitted to rotate and translate as whole entities and this was controlled by the intermolecular interactions being described using CSFF. These were referred to as the RBM MD simulations.
2. The cage atoms were described explicitly using CSFF, so that their flexibility could be included. These were referred to as the FBM MD simulations.

These simulations conditions were used throughout, unless otherwise stated.

The thermodynamic properties of the noble gases were also examined. For this, the Sorption tool within Accelyrs' Materials Studio[24] was used to simulate adsorption isotherms. The gas molecules were randomly inserted, translated and deleted within the **CC3** system, using the Metropolis Monte Carlo (MC) method. This treated **CC3** as a rigid body. For each gas, single component isotherms were

generated. The pressure range was 0.01 kPa to 1000 kPa, and 50 exponentially spaced fugacity steps were measured. The properties obtained for each sorption site were calculated every 50 steps, with a grid interval of 0.4 Å. After this, dual component adsorption isotherms were simulated with equal probability of loading krypton or xenon. This has a pressure range of 10 to 1000 kPa, with 50 exponential fugacity steps and a temperature of 273 K. 1,000,000 equilibration steps were used with a production run of 10,000,000 steps. The aim was to calculate the competitive sorption ratio of krypton and xenon.

5.4. Diffusion Coefficients

Diffusion rates are a vital tool for rationalising kinetic selectivity in porous materials. As the gases were tightly confined inside the pore of the **CC3** system, their diffusion rates were strongly affected by the shape and size of the pores and therefore this should give an indication to the influence of host flexibility, as has been seen previously for zeolites.[25,26] The self-diffusivity, $D_s(c)$, measures the displacement of a molecule as it diffuses through the system at equilibrium conditions. This was defined using the Einstein expression:

$$D_s(c) = \lim_{t\to\infty} \frac{1}{6t} \left\langle \left| \vec{r}(t) - \vec{r}(0) \right|^2 \right\rangle \qquad \text{(Eq. 5.1)}$$

Where $D_s(c)$ was the self-diffusivity, *t* the time, *r(t)* was the position vector of the diffusing gas at time *t* and $\vec{r}(t)$ was the distance travelled by the diffusing gas over a time interval of *t*. This was then averaged, indicated by the angular brackets.[25,27-29]

As the self-diffusivity can theoretically only be calculated accurately when *t* tends to ∞, the longer the simulation, the more accurate the results and hence why longer production runs were essential. Therefore, these simulations were run for 10 ns, with one gas molecule within each system; a frame was output every 1.0 ps so that the MSD could be calculated. These results could then be compared to literature data, and subsequently the relative kinetics of the gases understood. When looking at the MSD as a function of time, it was evident that ballistic motion was seen for the start of the simulation; this has been explained in more detail in Section 3.10. Therefore, a linear section of each MSD plot, where the slope was close to 1.0 on the logarithmic plots, was taken to represent the self-diffusivity of the gases. Examples of other porous materials and the diffusivity of

gases through them can be found in Appendix A.6. It might be expected that **CC3** would show broadly similar kinetic properties for the gases.

5.5. Screening criteria

To determine how useful these simulations are, a set of screening criteria were applied to the simulations, so that their properties could be identified and subsequently compared. The screening criteria measured were as follows:

1. The pore limiting diameter (PLD) for the RBM MD simulations.
2. The pore limiting envelope (PLE) for the FBM MD simulations.
3. The size of the gases used.
4. The effect of dynamic connectivity.
5. Shape exclusivity of the gases.

Three terms were important when analysing the results of the MD simulations; the pore aperture, the pore limiting diameter (PLD), and the pore limiting envelope (PLE). The PLD was the minimum diameter along the largest interconnected path through a periodic material; this was discussed in more detail in Section 5.5.1. This can be used to screen RBM MD simulations as the restriction of the intramolecular forces makes this the limiting factor for pore connectivity throughout the calculation. The PLE was used to screen systems where FBM MD simulations were used. Here the PLD was measured as a function of time during the MD simulation, and subsequently plotted as a histogram. More details on the PLE can be found in Section 5.5.3. By using these measurements it was possible to ascertain how much the flexibility of the system impacted the diffusion of the gases as well as looking at their kinetic properties. In addition, the thermodynamic properties of the noble gases were examined.

5.5.1. Pore Limiting Diameter (PLD)

The PLD has previously been used to explain barriers to diffusion, and was calculated using van der Waals radii for the framework atoms in accordance with the procedure developed by Haldoupis *et al*.[8,25,30] This value can be used to screen potential gas targets by plotting the kinetic diameter of desired molecules against the PLD to see whether the gases could potentially diffuse through the system. If these targets have a smaller kinetic radius than the PLD, then, in theory, they should be able to move through the porous systems with little difficulty. Kinetically, the diffusion rates of these gases will decrease as the size of the gas and the PLD tend towards equivalence. This diffusion rate will decrease

until the PLD becomes smaller than the kinetic diameter of the gas; at this point the gas would no longer diffuse through the system.

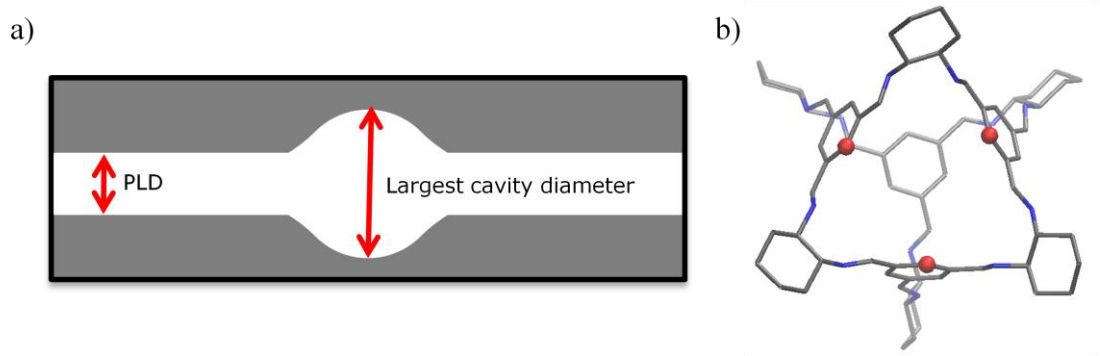


Figure 5.4 – Figure explaining the difference between the pore limiting diameter and the largest cavity diameter. a) Figure adapted from Haldoupis *et al*[8,31,32] which illustrates a two-dimensional representation of a porous material. The PLD was shown as the narrowest diameter of the pore channel, while the largest cavity diameter was shown as the widest point of the connected pore network. b) The location of the PLD for **CC3** was limited by the three phenyl carbons situated around the edge of the window. Nitrogens are blue, carbons grey and all hydrogens have been omitted.

To test whether the PLD was a useful measure to use for **CC3**, the diameter of the gases were compared to the PLD for **CC3**. First though, the PLD for **CC3** needs to be established.

5.5.2. Measuring the PLD for CC3

The narrowest point of the pore topology for **CC3** was located at the point where a window was formed within each discrete cage; the size of this window was determined by the phenyl carbon atoms pointing into their respective cage windows. To represent this, a circumcircle was used to define a geometric feature representative of the window. A circumcircle, otherwise known as a circumscribed circle, was a circle that passes through all the vertices of a stated polygon. Here, this was drawn between the three phenyl carbons located for each open cage face and hence represents the circle generated by a triangle and is shown in Figure 5.4b. So that the three corresponding carbon atoms for each window can be identified, the C-H bond vectors between the phenyl carbon and the attached hydrogen atom were used. As the three atoms in the same window should all be point in the same direction (out of the window), the dot product of any pair of C-H bond vectors within a cage will give a positive result. However, other C-H bond pairs also generate positive vectors and therefore a secondary check was put into place. This ensures that the correct number of intervening bonds between the carbon atoms was correct; this should be nine. The combination of the two tests

enables the identification of the correct sets of the three C atoms that define each cage window and hence define a circumcircle representative of the window size.

The centre of a circumcircle for a triangle was at the intercept of the three perpendicular bisectors of the sides; this point was known as the circumcentre. It can be found from the atom co-ordinates using the following geometric construct. Firstly, the atom co-ordinate position vectors a_i were used to define vectors for the three sides of the triangle, s_{ij} given by:

$$s_{ij} = a_j - a_i \qquad \text{(Eq. 5.2)}$$

In Figure 5.5 the order of the atoms were arranged so that the s_{ij}-vectors run clockwise around the triangle. The directions of the perpendicular bisectors, b_{ij}, were normal to the sides and in the plane of the triangle, *i.e.* perpendicular to the plane normal, n, which can be found from the cross product of any pair of sides. The vectors of the side bisectors were then:

$$b_{ij} = s_{ij} \times n \qquad \text{(Eq. 5.3)}$$

The position vectors of the side centres, m_{ij}, can be obtained by adding half the appropriate s_{ij} vector to each atom position vector. The b_{ij} were set to unit vectors that point along the line of the perpendicular bisectors toward the centre of the triangle. For any point along the line of a particular bisector we can then write:

$$v_{ij} = m_{ij} + \lambda_i b_{ij} \qquad \text{(Eq. 5.4)}$$

Where λ_i was a scalar multiplier. At the circumcentre the three bisectors meet and so any pair of vectors can be used to find the value of one of the λ_i. For example from sides 12 and 23:

$$v_{12} = m_{12} + \lambda_1 b_{12} \quad \text{and} \quad v_{23} = m_{23} + \lambda_2 b_{23} \qquad \text{(Eq. 5.5)}$$

This means that:

$$m_{12} + \lambda_1 b_{12} = m_{23} + \lambda_2 b_{23} \quad \text{rearranges to:} \quad \lambda_1 b_{12} - \lambda_2 b_{23} = m_{23} - m_{12} \qquad \text{(Eq. 5.6)}$$

It was then possible to obtain a solution from simultaneous equations formed from the dot products:

$$\lambda_1 - \lambda_2 b_{12}.b_{23} = b_{12}.(m_{23} - m_{12}) \text{ and } \lambda_1 b_{23}.b_{12} - \lambda_2 = b_{23}.(m_{23} - m_{12}) \qquad \text{(Eq. 5.7)}$$

Where we have used the fact that the b vectors were unit.

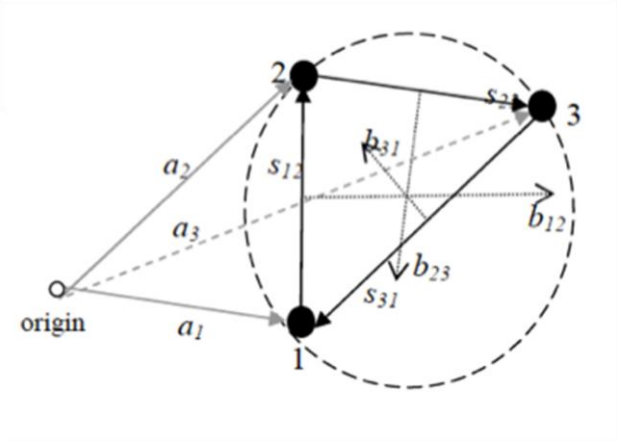


Figure 5.5 - The geometric construct used for defining the circumcircle from three atom co-ordinates (1,2,3). The centre of the circle passes through all three atoms and was found from the intercept of the perpendicular bisectors of the three sides (b_{12}, b_{23}, b_{31}).

Using $b_{12}.b_{23}$ as a multiplier for the second equation in (Eq. 5.7) it was possible to eliminate λ_2 to obtain:

$$\lambda_1 = \frac{(b_{12} - (b_{12}.b_{23})b_{23}).(m_{23} - m_{12})}{1 - (b_{12}.b_{23})^2} \qquad \text{(Eq. 5.8)}$$

Finally, using this value the vector for the centre of the circumcircle can be found from (Eq. 5.4). The actual circumcircle calculated was overestimated, as the van der Waals radii of the phenyl carbons has not been taken into account.* Therefore to calculate the PLD, the van der Waal radius of a carbon atom[19] (1.70 Å) was taken away from the calculated circumcircle radius; this was then doubled to give the PLD.

5.5.3. Pore Limiting Envelope (PLE)

When the flexibility of the cages is considered, the assumption of using the traditional PLD would prove a restriction, because this was taken from the static structure and does not take flexibility into account. Monitoring the PLD as a function of time, and subsequently plotting the frequency of these values in the form of a histogram, allows the flexibility to be taken into account in the form of a 'pore limiting envelope' (PLE). To calculate an example of a PLE, the guest-free **CC3** system simulated using the FBM MD simulation conditions discussed above.

* This was calculated using analyse_hist, a script developed by Dr David Willock, University of Cardiff.

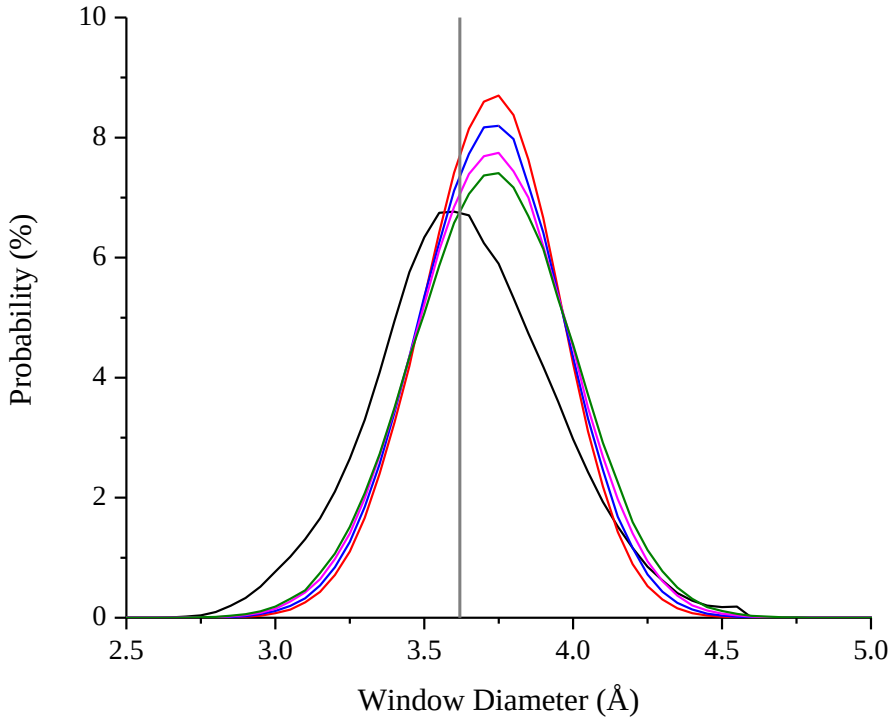


Figure 5.6 – Figure showing the PLE for the empty, guest-free **CC3** system during a FBM MD simulation at varying temperatures; 77 K (black), 250 K (red), 298 K (blue), 350 K (pink) and 400 K (green). The solid grey line represents the PLD (3.62 Å) from the crystal structure.

Geometric analysis of the trajectory files produced from the MD run was carried out using an in-house code that generated the circumcircle for each frame and hence a PLE for the cage window. A histogram of PLD against frequency was plotted to show the PLE, as shown in Figure 5.6. The PLE may be affected by the differing degrees of motion associated with different temperatures and therefore the effect of altering the temperature of the simulation on the distribution of the PLE needs to be understood. If this effect is strong, a single temperature PLE may not be an ideal screening tool, the results would be severely compromised by any temperature change. Using the same conditions as in Section 5.3, MD simulations were run at temperatures of 77, 250, 298, 350 and 400 K. The PLE was calculated for each simulation temperature; Figure 5.6. The results show that the maximal window diameter for **CC3** was around 5 Å for all temperature ranges and that, although the 77 K distribution was shifted slightly to the left, the others, 250, 298, 350 and 400 K, all behave in a similar manner. This suggests that the PLE of the empty **CC3** cannot exceed the maximal value of 5.0 Å and therefore only the distribution of the histogram was affected. As temperature was increased, the probability for the window diameter to be larger, with respect to the histogram at 298 K, increases. Additionally, a minimum window diameter of around 2.70 Å was observed. The 77 K distribution has practical relevance since this is the temperature at which nitrogen sorption was performed.

When comparing the PLD to the PLE, Figure 5.6, it was evident that the PLD was very close to the maxima of the PLE histogram. This suggests that the range of values observed for the PLD during the MD simulation oscillates around the static PLD value. This means that the PLD was smaller than the static PLD at times during the FBM MD simulation, and at other times it was larger; the probability for this was determined by the frequency of the histogram. One of the advantages with using the PLE instead of the PLD was that rigid host approximation is removed. This may not be important when large pores are considered, but when the size of the PLD tends towards the size of the gas, the use of the explicit value for the PLD will prove a limitation and lead to inaccurate conclusions. Additionally, when using the PLE there was both a maximum and minimum value; these can be used to more precisely screen systems for a given gas separation.

5.5.4. Sizes of the gases

One way to predict valuable traits of porous host systems, is to set screening criteria. One such screening criterion was the size of the molecules compared to the host structure. To generate these, the radii of the atoms needed to be decided upon. Throughout the literature, there are a whole host of different values for the radii of the different atoms, suggesting that the actual values are difficult to define. For example, the diameter of krypton has been reported as 3.60,[16,33] 3.636[16,34,35] and 4.04 Å.[19,34] Therefore, for consistency, the radii within this work was taken from the Cambridge Crystallographic Data Centre CCDC,[19,33] with the exception of both krypton and xenon; these were taken directly out of the recent work published by Sikora *et al*.[1,2,18] This has allowed us to determine explicit values for the radii of the gases that can be used to screen **CC3**; for example, the linear gases, hydrogen, nitrogen and carbon dioxide, can be considered at their smallest when an end-on profile was considered. Full details of how these sizes were calculated can be found in Section 3.10. It was evident that hydrogen was by far the smallest at 1.09 Å, while sulfur hexafluoride was the largest gas considered here at 6.05 Å. Sulfur hexafluoride was included to act as a control, as it was unlikely to diffuse through the **CC3** solid due to its size. These are summerised in Table 5.1.

Gas	Diameter (Å)
Hydrogen, H_2	1.090
Nitrogen, N_2	1.550
Carbon Dioxide, CO_2	1.700
Methane, CH_4	2.126
Krypton, Kr	1.818
Xenon, Xe	2.050
Sulfur Hexafluoride, SF_6	3.027

Table 5.1 – Table showing the van der Waal radii of the gases.

5.5.5. Different types of diffusion

Throughout this work, three different types of diffusion will be discussed; passive, opportunistic, and cooperative. These are defined as follows:

- **Passive diffusion** is when the gas can diffuse freely throughout the system without having an effect on the host structure; both the host and gas are independent. This is likely to be observed when the cages in the host system are treated as rigid bodies. Hence while this keeps the PLD constant, the gas should diffuse through the system unhindered. Passive diffusion is likely when the pores are substantially larger than the gas. Crucially, the orientation of the gas with respect to the pore is.

- **Opportunistic diffusion** is when the guest either takes advantage of a change in the host structure, or has to re-orient itself to fit through the narrowest point in the pore topology: it cannot passively diffuse through the system. The gas should therefore be able to move through both the rigid body and fully flexible simulations, but not without either re-orientation of the host or the guest. The trajectories of the gases must be looked at in detail to see whether their pathways are restricted by the system, and hence whether the diffusion is 'opportunistic', rather than passive.

- **Cooperative diffusion** is when the guest only moves through a system when the host structure accommodates such a movement by changing shape or orientation. This means that this diffusion would only be seen when the host system was considered fully flexible, as the host cannot accommodate the gases if they were kept rigid.

Distinguishing between these types of diffusion was possible by comparing the trajectories of the gases in both the RBM and FBM MD simulations. The RBM simulations can be used to distinguish between passive and opportunistic diffusion; if the gas diffuses without having to re-orientate, then the diffusion type was passive. If the gas has to re-orientates to fit through the window, then

the diffusion was more opportunistic as the gas was taking advantage of the space *via* re-orientation. This can be distinguished by looking at the trajectories of the gases whilst it transverse the cage window.

The FBM MD simulations can be used to distinguish between opportunistic and cooperative diffusion. If the gas moves through the system, but does not directly affect the host lattice in any way, then opportunistic diffusion was being observed. This can be determined by looking at the size of the cage window as the gas approaches; if it was unaffected opportunistic diffusion was taking place, as the cage window does not have to alter from the norm as the gas traverses it. Alternatively, if the guest invokes an abnormal response in the host structure, cooperative diffusion was taking place. This can be a change in the breathing motion of the cage window, or an unexpected shift in the PLD. This was because the gas was forcing the system to accommodate a change to allow the diffusion. This can also be ascertained by asking two questions:

1. What happens to the PLE as the guest approaches the window?
2. Does the window diameter behave abnormally when the gas transverses the window?

5.5.6. Comparison of the gas size to both the PLD and PLE

The window diameter of **CC3** was calculated to be 3.62 Å; this corresponds to the PLD. This can be used as a screening variable by comparison to the diameter of the gas molecule, as shown in Figure 5.7. Interestingly, this simple analysis shows that **CC3** should be almost ideal for krypton/xenon separation, as xenon was larger than the PLD and therefore unlikely to diffuse, whereas krypton was on the cusp of the PLD and should, in theory, diffuse through the system.

By allowing for flexibility, the value for the window diameter changes frame to frame during the FBM MD simulation. This means that there is now a range of values, with their frequency indicating the likelihood of the window diameter being open at this diameter. By using the PLE, the maximal window diameter was increased by around 1.5 Å (Figure 5.8). This has a large effect on the assessment of which gases will be able to diffuse through the cage system, as now only sulfur hexafluoride looks to be restricted. This would suggests that if the PLE was used as a screening mechanism, **CC3** might not be a 'perfect' candidate for separation of krypton from xenon, as both should, in theory, diffuse.

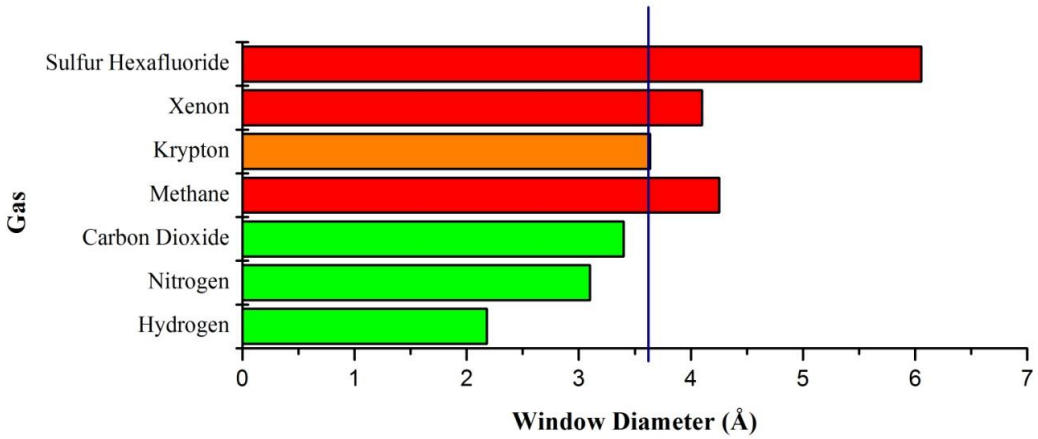


Figure 5.7 – The PLD, represented by the blue line, with respect to the diameter of the various gases. Green bars indicate the gases that would be expected to diffuse through the system, as they are smaller than the PLD; red bars indicate the gases restricted by the window diameter.

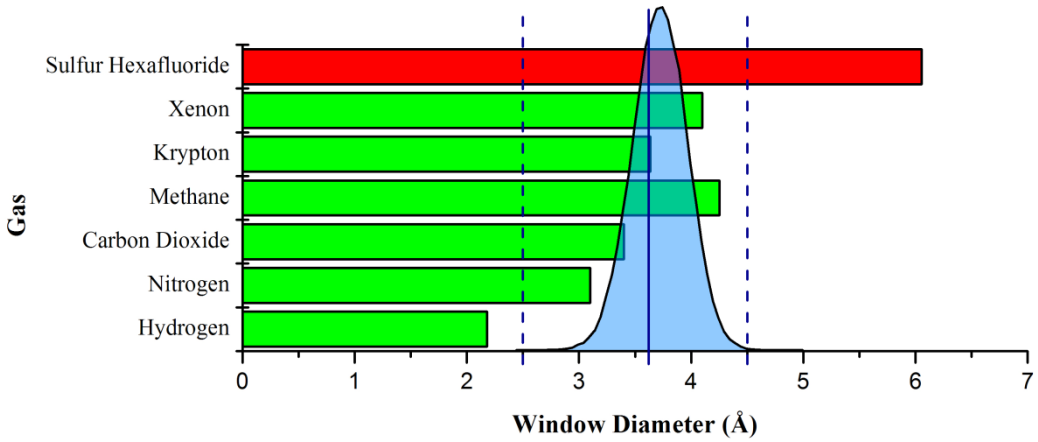


Figure 5.8 – Graph showing the PLE versus the size of the different gases. It was clear that opposed to a single value for the window diameter (solid line), the PLD, the PLE generates a minimum and maximum (dashed lines). This increase maximum amplifies the potential for larger gases to pass through the system. Green bars indicate the gases that would be expected to diffuse through the system; red bars indicate the gases restricted by the window diameter.

5.5.7. Does the gas proximity affect the PLE?

When looking at the FBM MD simulation data, it was difficult to distinguish between opportunistic and cooperative diffusion. One approach was to generate the PLE as a function of the proximity of the gas; this should help to ascertain whether the behaviour of the PLE was altered when the gas was in close proximity. This can then be compared to the empty **CC3** system as a reference. Figure 5.9 shows how the proximity of the gas was measured. A sphere from the centre of the cage window was constructed with a radius, τ_0. This was used to select the data with which the PLE was generated. If cooperative diffusion was

taking place, the PLE would be affected to a greater extent as τ_0 tends towards zero. It was difficult to ascertain whether the gas diffusion was opportunistic (the gas takes advantage of the wider window to move through the system), or whether the system was accommodating diffusion by responding to the close proximity of a guest by widening the window. That is, a correlation between gas approach and window opening is not necessarily 'causal' in nature. Another way to see whether cooperative diffusion was taking place was to plot the fluctuation in size of the window diameter away from the PLD as a function of time, and compare this to the value of τ_0; if the fluctuation of the window diameter alters as τ_0 crosses the origin, this would suggest that the window was being affected by the guest, thus suggesting cooperative diffusion. If the fluctuation was unaffected, then opportunistic diffusion was occurring. This distinction was important, as it shows whether guests within **CC3** affect the host structure; this was a vital consideration when analysing a candidate for gas separation.

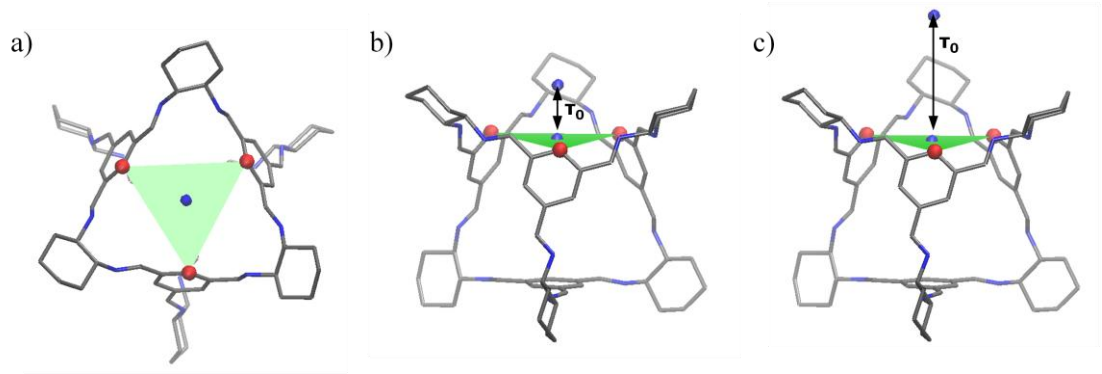


Figure 5.9 – Diagram showing that τ_0 was the distance between the circumcentre, shown in a) and the gas molecule shown in b) and c). As τ_0 was increased, it moved further away from the cage window. The green triangle represents the position of the cage window and the red spheres denote the three phenyl carbons used to describe the circumcircle. All hydrogens are omitted for clarity.

5.5.8. Dynamic Connectivity

The limitation of using the static X-ray crystal structure as a representative for the **CC3** system is that it assumes that flexibility is not important. For porous materials with larger pore channels this assumption may well be valid, but **CC3** has a narrow pore topology, and therefore to ignore its flexibility was unreasonable; this could miss out such phenomena as dynamic connectivity. Dynamic connectivity is where the pore topology was disconnected for the static X-ray crystal structure, but becomes connected when the superimposition of the frames during a MD simulation were looked at. The concept here was that different points of the pore topology were open at different time periods and

hence although a channel may be disconnected for a snapshot in time, overall there was a transient channel present which helps facilitate gas diffusion; this idea is illustrated in Figure 5.10.

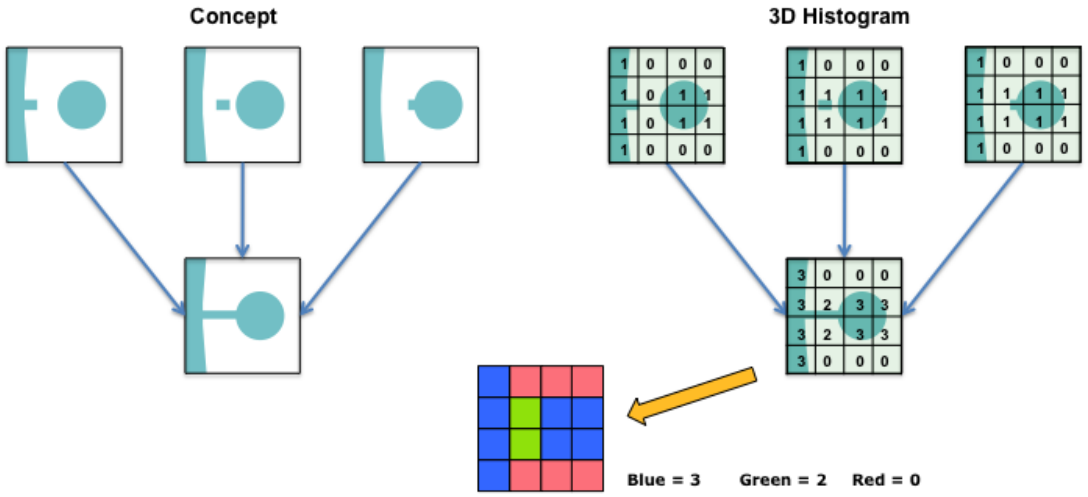


Figure 5.10 – Scheme showing the concept behind dynamic connectivity. Here each step of the MD simulation was divided into separate images and these were split into a grid. Each section of this grid was then given a number; 1 if it was in the pore structure, and 0 if it was not. These values were then summed over the whole simulation, and subsequently colour coded dependent on the values obtained, so that the pore connectivity can be highlighted by how long the pores were opened during the simulation.

By splitting the system into a grid, and then sampling whether each section was within the pore structure or not, a binary cube file can be generated, which can be used to generate a 3D histogram. This allows each part of the pore topology to be given a value and therefore how often each section of the connectivity was "open" can be shown. RBM and the FBM MD simulations of an empty **CC3** system were used, with a shorter production time of 1 ns. The limitations on the pore topology were dictated by the PLD and the PLE respectively; this should give a good indication of whether or not the increased flexibility made a large difference. Therefore, the radii of the different gases were used as the probe radius in the calculations and the results of the empty MD simulations compared. The results were coloured dependent on how long the pore topology was considered "open" during the 1 ns simulation; red, green and blue indicate times of 1, 10 and 100 ps respectively. "Open" corresponds to the area in question being accessible to the size of the probe used. This work was done in collaboration with Dr Maciej Haranczyk from the Lawrence Berkeley National Laboratory.[*] It was evident that the description used to describe the PLD as a circle was not sufficient in all cases.

[*] The source code for zeo++ was adapted to generate the 3D histograms.

This was because the actual shape of the window was more triangular in nature (Figure 5.11). This means that if the gas were to re-orientate, it may be able to pass through the window, even though it was described as being too big. To address this, the shapes of the gases were compared to the window of a single cage to bring to light any shape effects as a result of the linear, spherical tetrahedral and octahedral representations of the gas molecules.

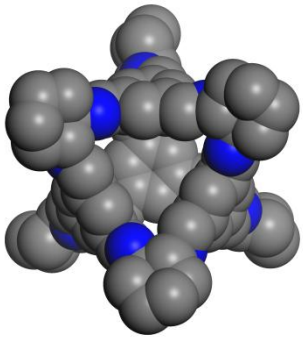


Figure 5.11 - This illustrates that the **CC3** cage window was triangular in nature. The hydrogens have been omitted for clarity.

5.6. Hypothesis

When you compare the PLD of **CC3** to the size of some of the gases, for example krypton and xenon, they were very similar; therefore it would be expected for the RBM MD simulations to differ to the FBM MD simulations. The PLE should be more useful when comparing with experimental sorption behaviour. If this proves to be correct, it would give a valid reason to include flexibility in future MD simulations as well as show how important the flexibility of these porous molecules was when consider prediction in the future.

5.7. MD simulations of 'linear' gases

The three 'linear' gases simulated were hydrogen, nitrogen, and carbon dioxide. Although they were all linear, they vary in size and therefore the kinetics of their diffusion and diffusion pathways could be different. Due to the charged nature of nitrogen and carbon dioxide, these were described as multipoles. To achieve this, nitrogen has a massless site to balance out the charge. Further details on gas fitting can be found in Section 3.10.

5.7.1. Hydrogen

For hydrogen, experimental sorption isotherms taken up to a pressure of 1 bar, and a temperature of 77 K, show an uptake of around 4.0 mmol g^{-1} (Chapter 1). The static pore topology also shows continuous 3D channels through the system (Figure 5.1a). It was important to see whether this was still the case during the

MD simulations, therefore the dynamic connectivity for hydrogen was examined for both the RBM and the FBM MD simulations, by screening them using a probe radius of 1.09 Å. It was clear that for hydrogen both simulations give a 3D pore topology. The results were shown in Figure 5.12. The red isosurface illustrates what parts of the system were considered accessible for 0.1% of the 1000 ps production run, the green isosurface illustrates where the pore channel was considered accessible for 1% of the simulation, whilst the blue isosurface indicates where the pore channel was open for 10% of the simulation; these have been split up in Figure 5.12 for clarity. The pore topology became more defined when the percentage of time that the pore remained open was increased.

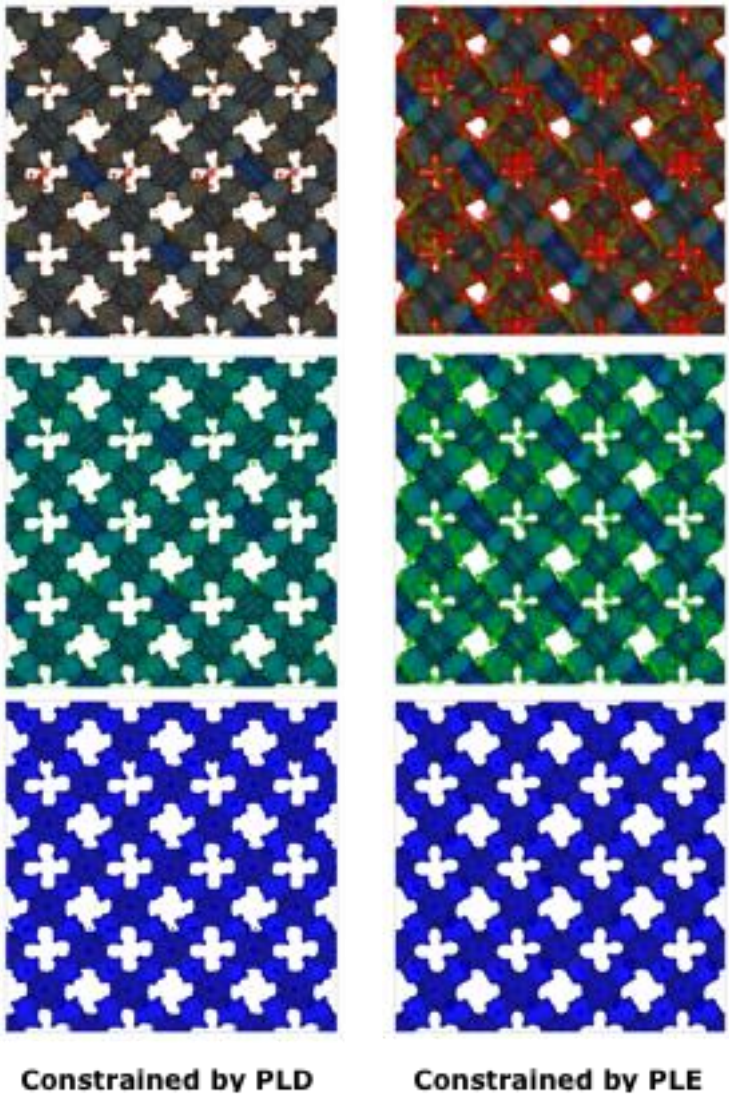


Figure 5.12 – Scheme showing dynamic connectivity. The dynamic connectivity for the RBM MD simulations is shown on the left, whilst the results for the FBM MD simulations are shown on the right. The pore was open for red = 1 ps (0.1%), green = 10 ps (1%) and blue = 100 ps (10%) of a 1000 ps production run. These have been split up for clarity.

This means that the blue isosurface shows where the gas is most likely to be, with the green and red isosurfaces indicating areas where the gas has less chance to be located; these sections show rather little difference in the case of hydrogen, but this becomes more important as the size of the gas is increased. There is minimal difference in the blue isosurface for the RBM and FBM simulations, and this suggests that the increased flexibility of the system would have little impact on the overall diffusivity of hydrogen. The difference in the green isosurface is more pronounced; suggesting a larger pore system for the FDM and this is seen more prominently for the red isosurface. By looking at the orientation of the gas in a cage window, Figure 5.13, it was evident that hydrogen should easily diffuse 'passively' through the system irrespective of its geometry, as it was far smaller than the pore aperture; the PLD of **CC3** (3.62 Å) was much larger than the maximum diameter used for the hydrogen, 2.18 Å. This suggests that the extra flexibility in the system would have little effect on the kinetic diffusivity and that there was no shape exclusivity taking place.

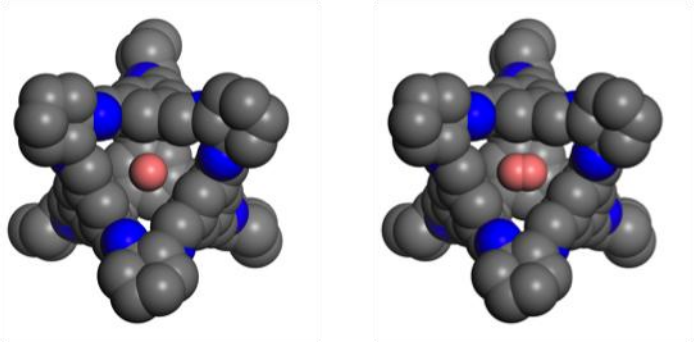


Figure 5.13- Scheme showing both the side-on and end-on profile of a hydrogen molecule (pink) within a single **CC3** window. All hydrogens on the cage have been omitted for clarity.

One molecule of hydrogen was then inserted into the centre of a cage within a 2 x 2 x 2 supercell of **CC3**. During the production run, an image was output every 0.1 ps; this allowed the coordinates of the hydrogen to be overlaid showing the pathway of the gas through **CC3**. During the 10 ns RBM MD simulations, the hydrogen diffuses into 100% of the cage cavities, and this can be illustrated by superimposing the gas position from each frame within the MD simulation, Figure 5.14a. The pathway of the hydrogen, Figure 5.14b, shows that the gas was freely diffusing through the system while Figure 5.14c confirms the original assumption that the gas does not need to re-orientate to pass through the cage window. This would suggest that using the RBM MD simulation to describe hydrogen might be acceptable, as the gas seems unaffected by the PLD as it diffuses through the system – that is, the diffusion is 'passive'.

The FBM MD simulations for hydrogen in **CC3** show a similar story; 100% of the cages were visited during the 10 ns simulation. Figure 5.14d was comparable to Figure 5.14a, as was the trajectory through a cage window, Figure 5.14b and Figure 5.14e. Figure 5.14f, shows that hydrogen was not being forced to re-orientate as it can pass from cage to cage right at the periphery of the cage window, indicating that the window was not affecting the diffusion; this was similar to the RBM MD simulations, Figure 5.14c. This resemblance was unsurprising, as the gas was very small and therefore the increased flexibility of the FBM MD simulations has little effect on the results, even though the connectivity and van der Waals interactions may be was enhanced. To examine this, the diffusion coefficients of both simulations were calculated; these were shown, along with their logarithmic plots, in Figure 5.15. Both of the diffusion coefficients were calculated from the normal diffusion range of the graph, where the slope of the logarithmic plot was close to 1.

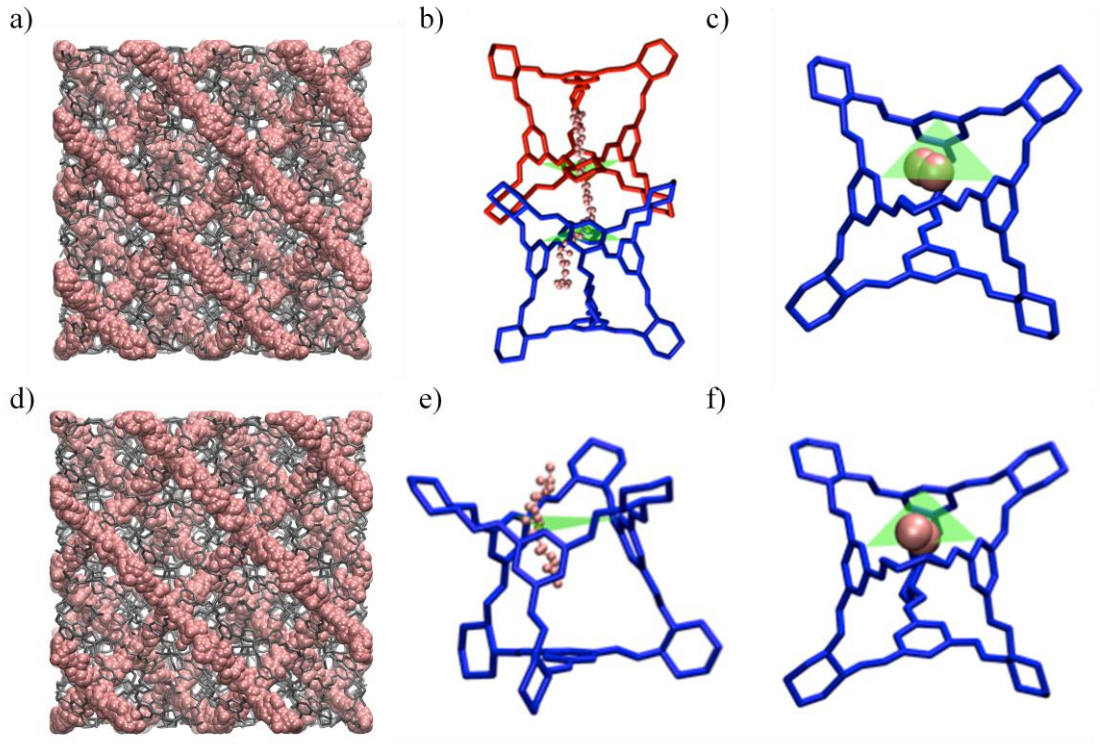


Figure 5.14 – This shows illustrations for RDM MD simulations of the a) position of H_2 at every point during the simulation, b) the trajectory of a H_2 molecule through the periphery of the cage window and c) its orientation of the gas as it diffuses through the window of **CC3**. For the FBM MD simulations illustrations of d) the position of H_2 at every point during the simulation, e) the trajectory of a H_2 molecule through the periphery of the cage window, and, f) its orientation of the gas as it diffuses through the window of **CC3**. H_2 has been coloured pink so that it was more visible within the figures, whilst all cages have been given a single colour, so that they stand out.

The diffusion coefficients for the RBM and FBM MD simulations were calculated to be 5.44 x 10^{-8} m^2 s^{-1} and 5.64 x 10^{-8} m^2 s^{-1} respectively. This suggests that the diffusivities of the gases were similar; meaning that little has been gained by running the FBM MD simulations. To validate this assertion, the PLE was measured, using τ_0 as a cut-off for the data, Figure 5.16a. The graph shows that the PLE for the system was not affected by any reduction in τ_0; the range, distribution and maxima of the PLE all remained the same as the reference **CC3** data, suggesting that hydrogen does not have an influence on the window diameter. Subsequently, the deviation from the static window diameter (3.62 Å) was measured as a function of time and plotted on the same axis as τ_0 versus time to see whether the proximity of the gas had any effect on the cage window, Figure 5.16b. The fluctuation of the window diameter stays well within the maximum and minimum values observed during the MD simulation, and no obvious change was seen as τ_0 tends towards zero, suggesting that the gas was having no effect on the dynamic motion of **CC3**. In summary, passive diffusion was observed, as **CC3** was not directly influencing the gas; hydrogen diffuses freely within the system.

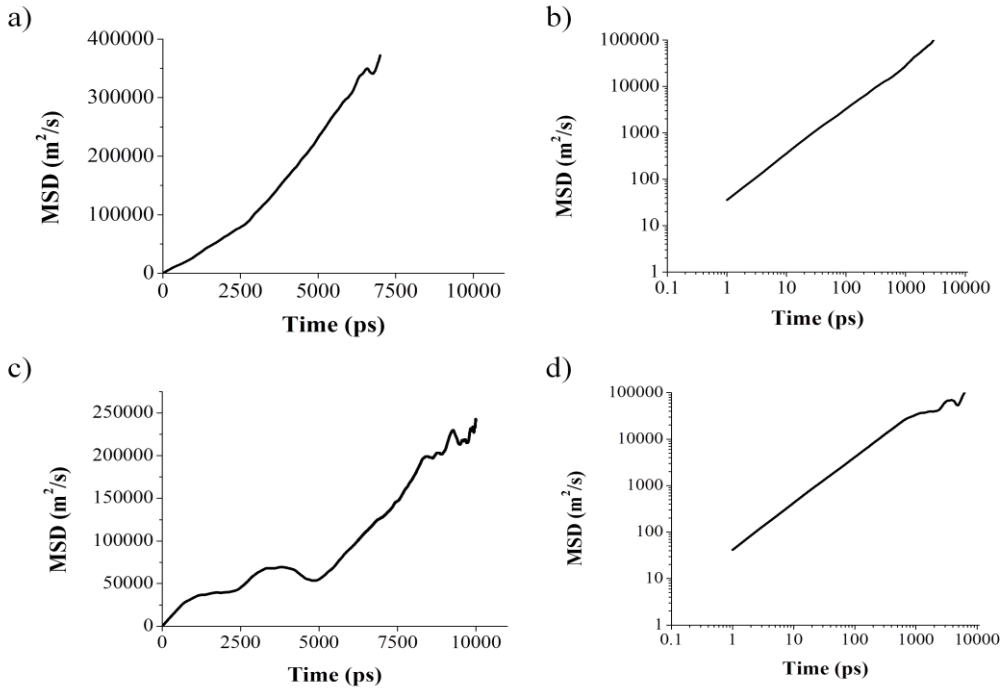


Figure 5.15 – The MSD for hydrogen in **CC3** for a) the RBM simulation, b) the logarithmic plot of this to check normal diffusion has occurred, c) the FBM MD simulation, and d) logarithmic plot for the FBM MD simulation.

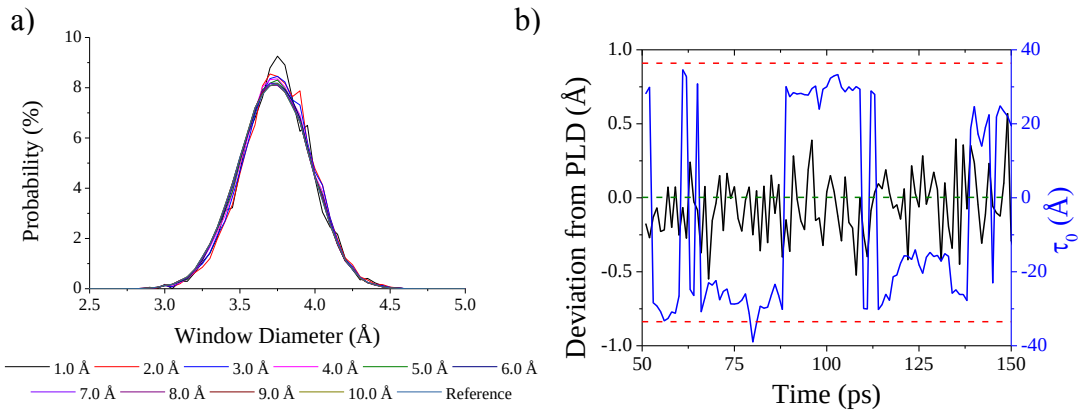


Figure 5.16 – Graph showing a) the PLE for a system containing hydrogen for decreasing values of τ_0 compared to the PLE of the empty **CC3** reference system and b) the deviation of the window diameter from the original (3.62 Å) as a function of time (black) plotted against τ_0 versus time (blue). The dashed red lines show the maximum and minimum deviations, whilst the origin was marked with a dashed green line.

5.7.2. Nitrogen

For nitrogen, experimental sorption isotherms taken at up to a relative pressure of 1 and a temperature of 77 K show an uptake of around 4.0 mmol g^{-1} (Chapter 1). As with hydrogen, the static pore topology showed continuous 3D channels throughout the system, though due to nitrogen being larger than hydrogen by 0.92 Å in diameter, the channels were slightly smaller (Figure 5.1b). The dynamic connectivity shows that although nitrogen is larger, there was a 3D connected pore topology for both the RBM and FBM MD simulations; Figure 5.17.

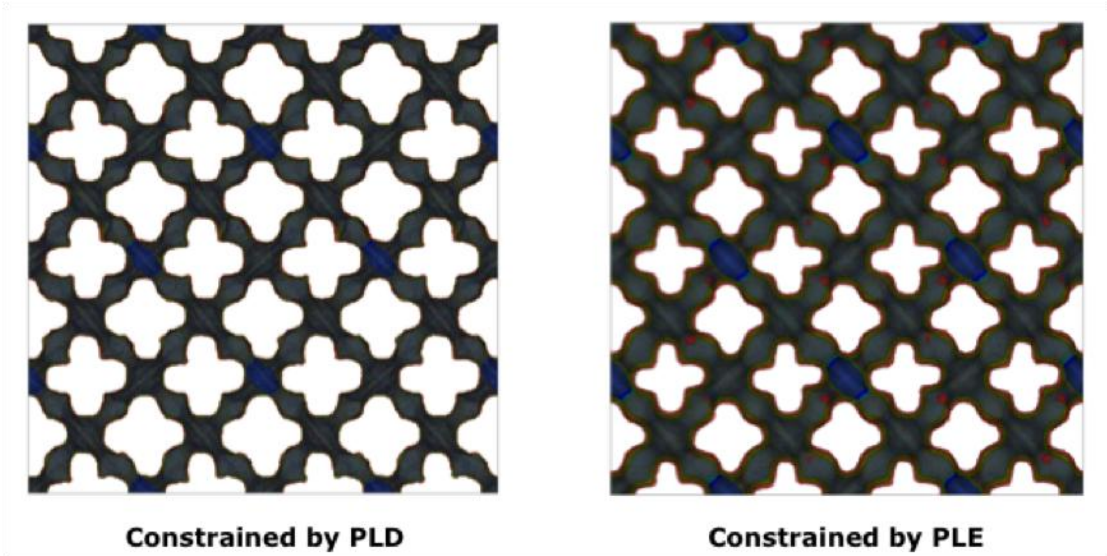


Figure 5.17 – Scheme showing dynamic connectivity for nitrogen, using a probe radius of 1.55 Å. The dynamic connectivity for the RBM MD simulations is shown on the left, whilst the results for the FBM MD simulations are shown on the right. The pore was open for red = 1 ps (0.1%), green = 10 ps (1%) and blue = 100 ps (10%) of a 1000 ps production run.

The results show that the simulations defined by the PLE show a much larger pore topology, and this was because of the increased flexibility of the system. This should have an effect on the diffusivity, enhancing it in the FBM simulation. The cage cavity is more pronounced in the FBM results, with the pathway from cage to cage, through the interstitial sites, larger in size; this is evident by the blue isosurface being more open in the FBM results and would suggest a larger pathway for the nitrogen to diffuse through. By placing a molecule of nitrogen in the centre of a cage window, Figure 5.18, it was evident that nitrogen should be able to fit through the window in both a side-on and end-on manner, though the latter was a tighter 'fit'. This may make re-orientation of the molecule into an end-on profile advantageous when it diffuses through the cage window.

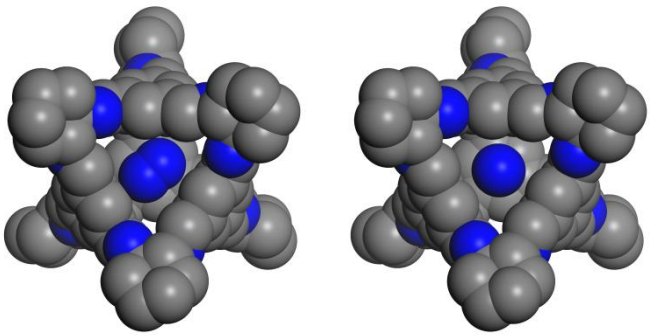


Figure 5.18 - Scheme showing both the side-on (left) and end-on (right) profile of a nitrogen molecule (blue) within the **CC3** window. All atoms were scaled to their CCDC radii and all hydrogens on the cage have been omitted for clarity.

To investigate the diffusivity of nitrogen, one molecule was inserted into the centre of a cage within a supercell of **CC3**. The results show that in the RBM MD simulations the nitrogen diffuses into 100% of the cage cavities over the 10 ns simulation and this is shown in Figure 5.19a – this was similar to hydrogen. This was unsurprising, as nitrogen can fit through the cage window regardless of its orientation. The diffusion pathway for nitrogen, Figure 5.19b, was unaffected as it moves through the cage window, showing that the gas was freely diffusing through the system. Figure 5.19c demonstrates that nitrogen can pass through the window with a side-on profile. The gas passes through the centre of the window. Like hydrogen, this suggests that using the RBM MD simulation to describe nitrogen might be sufficient in describing its diffusivity.

The FBM MD simulations for nitrogen in **CC3** allowed us to investigate if the increase in gas size affects the kinetic diffusivities. Over the 10 ns simulation 100% of the cages were visited. The superimposition of the trajectories, Figure 5.19d, was comparable to the RDM MD results, Figure 5.19a, whilst the orientation of the gas, Figure 5.19e, was also similar to its analogue, Figure

5.19c. What was interesting was that unlike the RBM MD simulation, nitrogen was now able to move through the cage window at its periphery, Figure 5.19f; this should increase the rate of diffusion, as the nitrogen is not restricted to moving through the centre of the cage window. The increased space available therefore increases the likelihood of diffusion, and consequently the kinetic diffusivity of the gas.

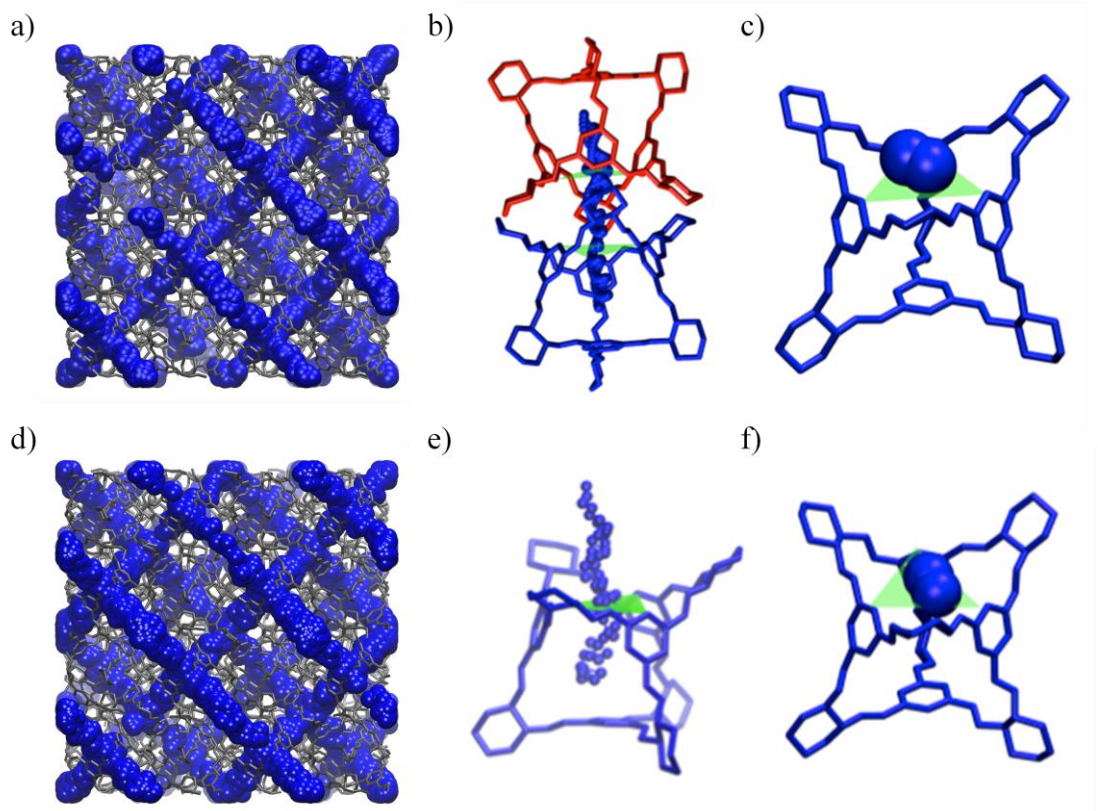


Figure 5.19 – This shows illustrations for RDM MD simulations of the a) position of N_2 at every point during the simulation, b) the trajectory of a N_2 molecule through the periphery of the cage window and c) its orientation of the gas as it diffuses through the window of **CC3**. For the FBM MD simulations illustrations of d) the position of N_2 at every point during the simulation, e) the trajectory of a N_2 molecule through the periphery of the cage window and f) its orientation of the gas as it diffuses through the window of **CC3**. All cages have been given a single colour, so that they stand out.

The diffusion coefficients for the simulations were calculated when the logarithmic slope, Figure 5.20, was close to 1, as this ensures normal diffusion was occurring. These show that the FBM MD simulations has a faster diffusion rate, with a diffusion coefficient of 5.94×10^{-8} m^2 s^{-1}, compared to the slower diffusion coefficient of 4.98×10^{-9} m^2 s^{-1} for the RDM MD simulations. This suggests that it was the increased flexibility that allows the gas to diffuse to a greater proportion of the system. To see if this has any effect on the PLE, τ_0 was again used as a cut-off for the PLE; as it was reduced it was compared to empty **CC3**; results for this can be seen in Figure 5.21a. It was evident that as τ_0 was decreased, there

was a slight affect on the PLE. When τ_0 was limited to 2 Å there was a small shift in the distribution, so that there was an increased probability of the window diameter being larger; this was more noticeable when τ_0 was reduced to just 1 Å. This suggests that nitrogen may be influencing the normal PLE.

To see if the gas affects the window during the simulation, the deviation from the static window diameter (3.62 Å) was measured as a function of time and plotted on the same axis as τ_0 versus time, Figure 5.21b. As the fluctuation of the window diameter stays well within the maximum and minimum values observed during the MD simulation it was suggested that the gas was not having an impact on the window diameter. This means that the shift in the PLE in Figure 5.21a was more likely down to the gas taking advantage in the swelled cage window, opposed to actually forcing it open. This means that nitrogen was also diffusing passively through the system, though unlike hydrogen where the diffusivities could be calculated directly from the BDM MD simulations, the FBM MD simulations for nitrogen were useful as they show that increased flexibility has an impact on the kinetic diffusivity of the gas.

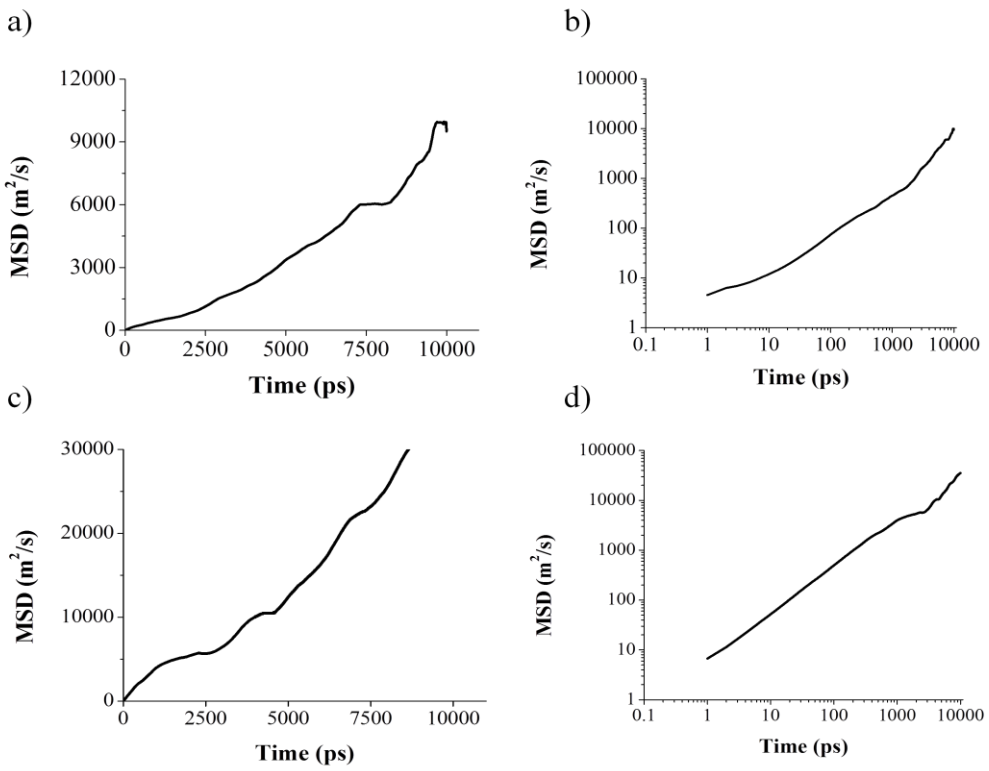


Figure 5.20 – The MSD for nitrogen in **CC3** for a) the RBM simulation, b) the logarithmic plot to confirm normal diffusion was occurring, c) the FBM MD simulation, and d) logarithmic plot for the FBM MD simulation.

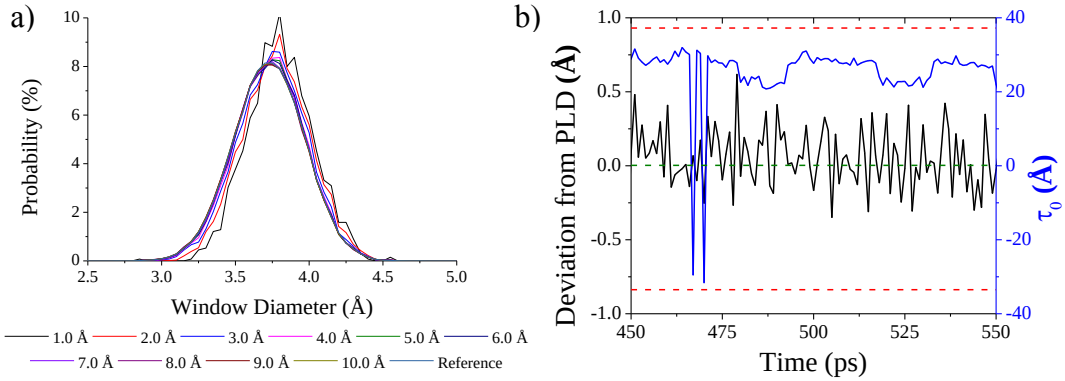


Figure 5.21 – Graph showing a) the PLE for a system containing nitrogen for decreasing values of τ_0 compared to the PLE of the empty **CC3** reference system and b) the deviation of the window diameter from the original (3.62 Å) as a function of time (black) plotted against τ_0 versus time (blue). The dashed red lines show the maximum and minimum deviations, whilst the origin was marked with a dashed green line.

5.7.3. Carbon dioxide

For carbon dioxide, experimental sorption isotherms recorded up to a pressure of 1 bar and a temperature of 77 K, show an uptake of around 2.0 mmol g^{-1} thus highlighting that carbon dioxide can diffuse through **CC3** (Chapter 1). This was about half the uptake seen for both hydrogen and nitrogen. A 3D diamondoid network is seen in the pore topology, Figure 5.1c, though these channels were much smaller as carbon dioxide was larger than both hydrogen and nitrogen (1.70 Å). Using this to calculate the dynamic connectivity, 3D pores were seen for both simulations, Figure 5.22. Similar to the nitrogen results, the simulations described by the PLE show a much larger pore topology, again suggesting that increased flexibility widens the pore channels. The RBM results for the dynamic connectivity show really narrow channels throughout the system, suggesting that the diffusivity for carbon dioxide would be slow. On the other hand, the channels for the FBM results were much more pronounced, therefore showing the propensity for the diffusivity of carbon dioxide to be higher. This would suggest that the increased flexibility in the FBM simulations is important and this can be confirmed by examining the diffusion coefficients.

For carbon dioxide, looking at the orientation of the gas in the window was vital, as Figure 5.23 suggests that only the end-on profile fits, therefore indicating that re-orientation would be necessary if the gas was to diffuse through the cage window. This suggests that shape exclusivity would take place during the MD simulations if diffusion was going to occur. To investigate the diffusivity of carbon dioxide, one molecule was inserted into the centre of a cage within a 2x2x2 supercell of **CC3**. During the RBM MD simulation, carbon dioxide covers less

volume than the previous two gases – only 58 % of the cage cavities were occupied during the simulation period, Figure 5.24a.

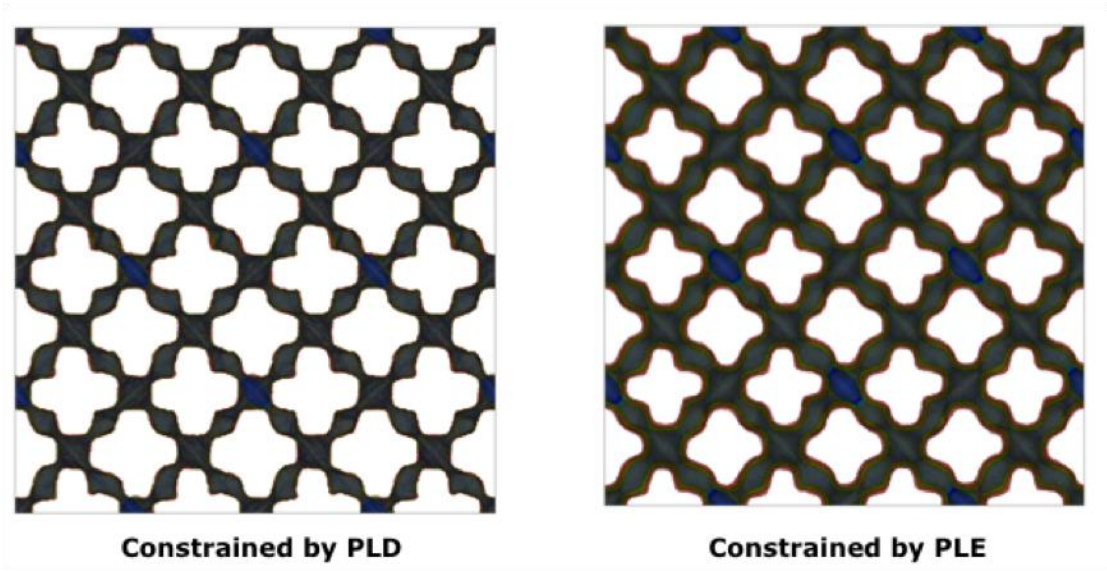


Figure 5.22 – Scheme showing dynamic connectivity for carbon dioxide, using a probe radius of 1.70 Å. The dynamic connectivity for the RBM MD simulations is shown on the left, whilst the results for the FBM MD simulations are shown on the right. The pore was open for red = 1 ps (0.1%), green = 10 ps (1%) and blue = 100 ps (10%) of a 1000 ps production run.

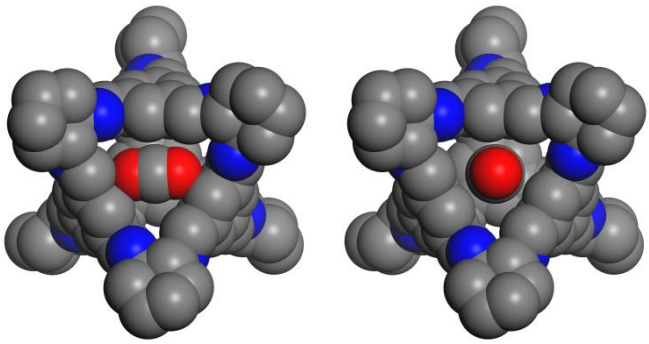


Figure 5.23 – Scheme showing both the side-on and end-on profile of carbon dioxide within the **CC3** window. All atoms were scaled to their CCDC radii and all hydrogens on the cage have been omitted for clarity.

As the molecule and pore were close in size, the gas tends to re-orientates to diffuse through the system; the smallest diameter of the carbon dioxide molecule was found when the molecule was viewed end on (3.40 Å), and this was close to, but still smaller than, the PLD of 3.62 Å. Figure 5.24b shows the trajectory of the gas as it moves from one cage to another. It was evident that carbon dioxide spends a long time in the interstitial site between the two cages – this was presumably due to the gas re-orientating so that it can move from the interstitial site to the second cage window. Figure 5.24c shows how the carbon dioxide re-

orientates to face the cage window end-on before diffusing through; this was observed for multiple windows.

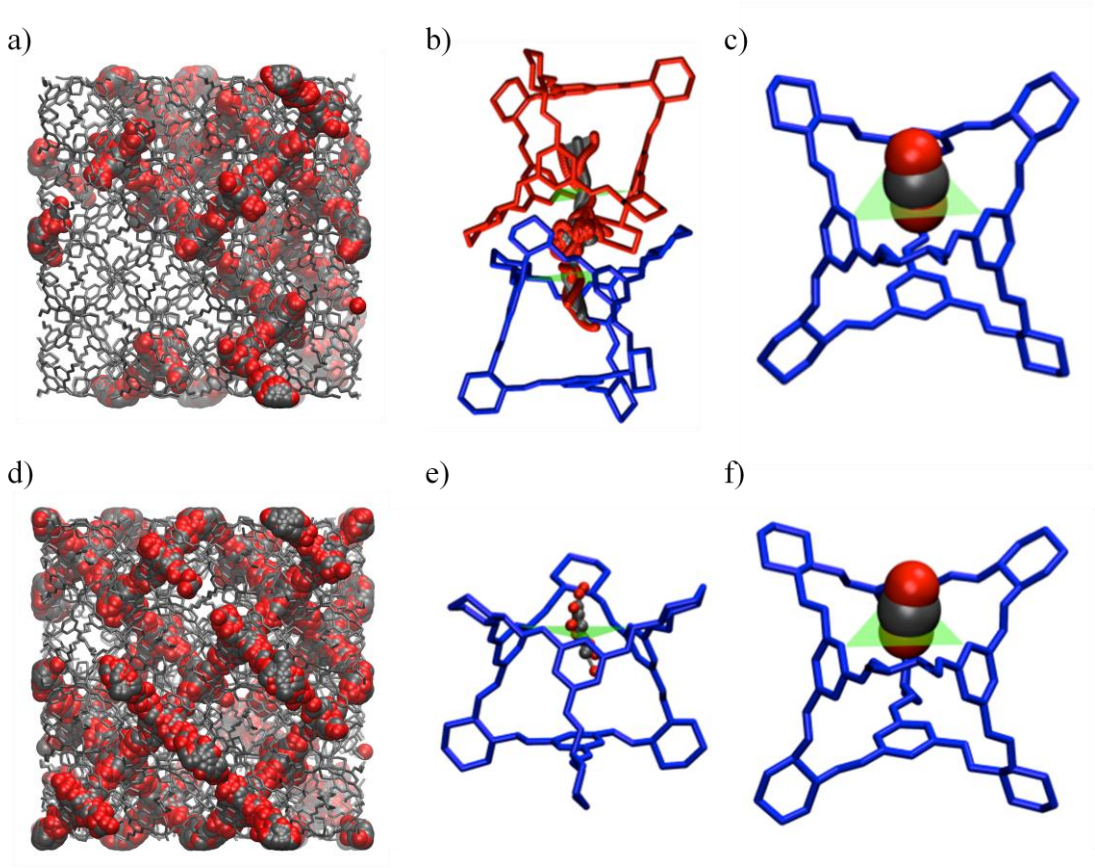


Figure 5.24 – This shows illustrations for RDM MD simulations of the a) position of CO_2 at every point during the simulation, b) the trajectory of a CO_2 molecule through the periphery of the cage window and c) its orientation of the gas as it diffuses through the window of **CC3**. For the FBM MD simulations illustrations of d) the position of CO_2 at every point during the simulation, e) the trajectory of a CO_2 molecule through the periphery of the cage window, and, f) its orientation of the gas as it diffuses through the window of **CC3**. All cages have been given a single colour, so that they stand out.

The re-orientation of the gas can be shown schematically by representing the carbon dioxide molecule as a dumbbell – it was then possible to watch this re-orientation as it diffuses through the cage window. The cartesian coordinates from the RBM MD simulations for the gas molecule were used as reference points for the dumbbell. Figure 5.25 shows that the gas molecule starts in the side-on configuration in the centre of the cage (red) and subsequently re-orientates to adopt the end-on profile as it moves through the window (pink). The re-orientation to the end-on profile reduces the size of the gas to 3.40 Å from 5.72 Å and subsequently carbon dioxide can then diffuse through the cage window. This was the first case of a gas not exhibiting passive diffusion within **CC3**. As the gas has to re-orientate to diffuse, the diffusion must be either opportunistic or

cooperative; this was investigated further with the more realistic FBM MD simulations.

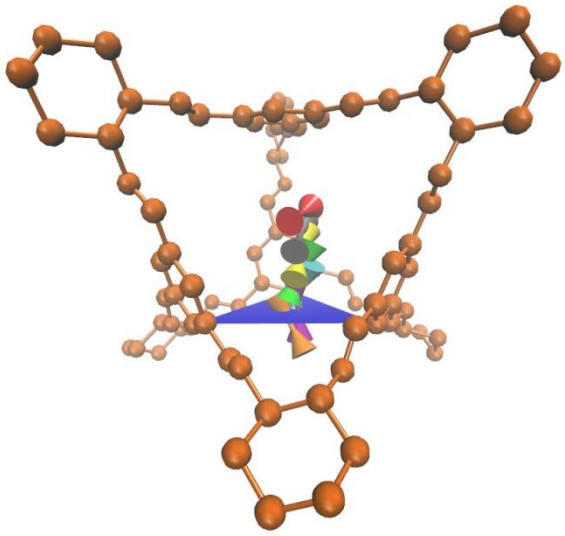


Figure 5.25 – Figure showing the re-orientation of a carbon dioxide molecule from a side-on profile to and end-on profile as if diffuses through the window of **CC3**. The gas has been drawn as a dumbbell for simplicity. The gas was coloured to show its position at different points in time; this proceeds red, grey, yellow, green, cyan, purple, and finally, orange.

The FBM MD simulations for carbon dioxide in **CC3** show that 63% of the cages were visited during the simulation; this was slightly more than in the RDM MD simulations (58%), and the overlay can be seen in Figure 5.24d. Figure 5.24e shows that to move through the cage window, the gas must still stay in an end-on conformation, otherwise it cannot diffuse; a more detailed view of this end-on orientation was seen in Figure 5.24f. The increased cage occupancy suggests that the FBM MD simulations should have faster diffusion rates. Indeed, the diffusion coefficient for the FBM MD simulation was two and a half times faster at 2.60 x 10^{-9} m^2 s^{-1}, opposed to the RBM analogue at just 9.72 x 10^{-10} m^2 s^{-1}. The relative diffusion coefficients suggest that the increased flexibility of the FBM MD simulations allow the gas to diffuse more quickly. This may be because the window of the cage was open to a greater extent in the FBM MD simulations, therefore facilitating the re-orientation of the gas. To examine whether cooperative diffusion was taking place, τ_0 was again used as a cut off for the PLE, Figure 5.27a. It was evident that as τ_0 was decreased, there was a very little effect on the PLE, therefore suggesting that the diffusion was more opportunistic in nature. To confirm that opportunistic diffusion was occurring, the deviation from the crystallographic window diameter (3.62 Å) was measured as a function of time and plotted on the same axis as τ_0 versus time, Figure 5.27b. As with both hydrogen and nitrogen, the fluctuation of the window diameter away from the original (3.62 Å) stays well within the maximum and minimum values,

suggesting that there was no impact on the window diameter as carbon dioxide travels through the window. This confirms that carbon dioxide was diffusing through the system without directly affecting the cage window. Due to its size the gas re-orientates and therefore this was an example of 'opportunistic' diffusion.

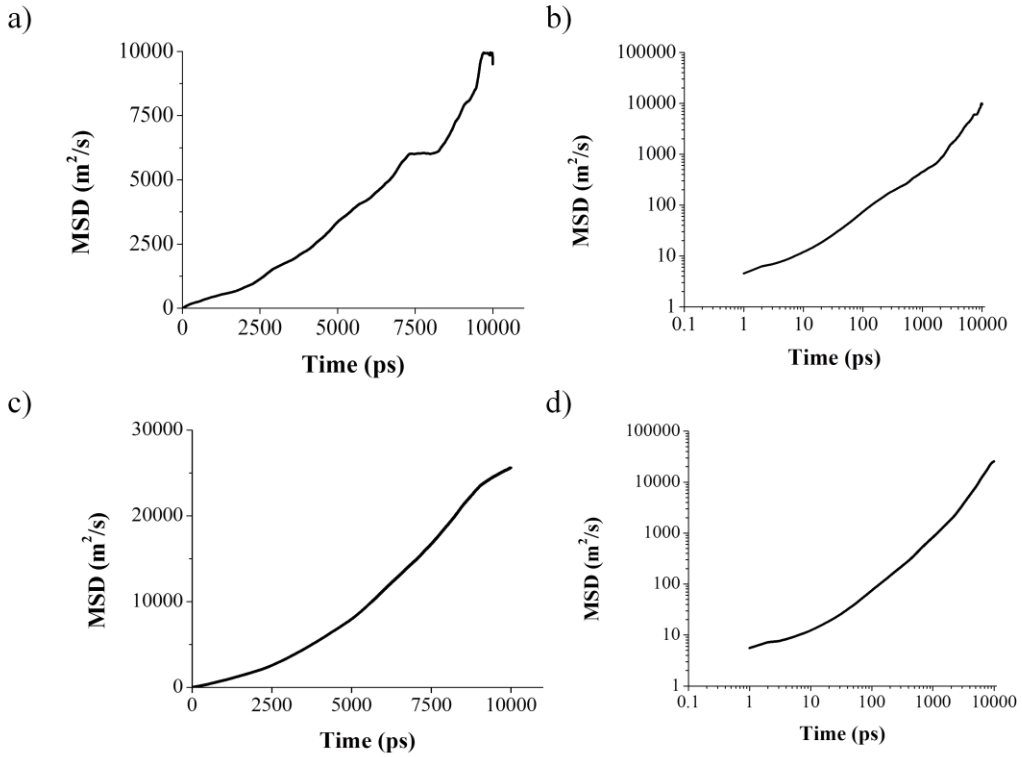


Figure 5.26 – The MSD for carbon dioxide in **CC3** for a) the RBM simulation, b) the logarithmic plot of this to check normal diffusion had been achieved, c) the FBM MD simulation, and d) logarithmic plot for the FBM MD simulation.

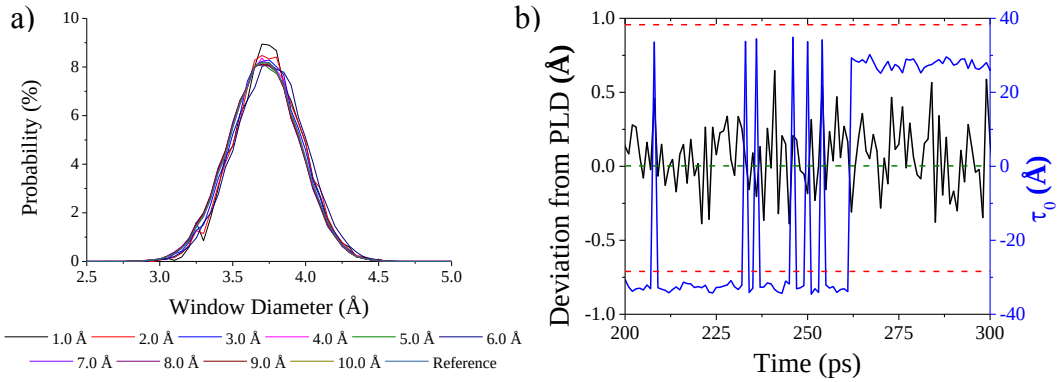


Figure 5.27 – Graph showing a) the PLE for a system containing carbon dioxide for decreasing values of τ_0 compared to the PLE of the empty **CC3** reference system and b) the deviation of the window diameter from the original (3.62 Å) as a function of time (black) plotted against τ_0 versus time (blue). The dashed red lines show the maximum and minimum deviations, while the origin was marked with a dashed green line.

5.8. MD simulations of tetrahedral and octahedral gases

Diffusion of one tetrahedral and one octahedral molecule were simulated within **CC3**. These were methane and sulfur hexafluoride, respectively. To see whether any kind of shape selectivity was taking place, each atom of the gas was described explicitly, as opposed to using a united atom approach.

5.8.1. Methane

For methane, experimental sorption isotherms show a similar uptake to carbon dioxide at around 2 mmol g^{-1}, at a temperature of 273 K and a pressure of 1 bar, (Chapter 1). The calculated pore topology, however, shows a disconnected structure, suggesting that methane would not be absorbed into **CC3**, Figure 5.1d. It was important to examine the dynamic connectivity carefully to see whether this was still the case. When described as a sphere, methane has a larger radius than all the three linear gases, 2.13 Å. This was evident when looking at the RBM MD simulation as the pores were limited by the PLD and this has an impact on the 3D pore topology; it was no longer connected and therefore diffusion would not be expected, Figure 5.28. Only small discrete pores were seen, one inside the cage cavity, the other in the interstitial site.

The FBM MD simulations show slight connections in the pore topology, therefore diffusion of methane could be envisaged, dependent on the length of the production run. The size within both the cage cavity and the interstitial site was increased in the FBM simulations, showing how important the flexibility was. A close up of the 3D pore topology for the FBM results is shown in Figure 5.29 showing partial connections. The blue isosurface shows isolated pockets, similar to those seen for the RBM simulations, but the increased flexibility means that channels could now be formed between these isolated voids. These channels are only seen in the red isosurface, suggesting that these are only open for small fractions of time within the simulation; the red isosurface only corresponds to the pore channel being open for 0.1% of the simulation time. In addition, only around 40% of these channels are transiently open through the simulation. The green isosurface displays a 'swelled' version of the blue isosurface, thus implying that the increased flexibility allowed the cage to swell and flex in nature. The interstitial sites were also more open for the FBM simulations, though this is unsurprising as this is directly linked to cage flexibility. It would be expected for methane to spend more time within the cage cavity opposed to the interstitial site, since these are more open.

This highlights a possible limitation in using our circumcircle methodology, because it assumes that all gases behave as spheres. Methane is, in fact, tetrahedral and if treated as such, the gas can re-orientate itself to fit into the cage window, Figure 5.30. This was possible because the window itself is triangular in nature; that is, the same shape as a face of a regular tetrahedron.

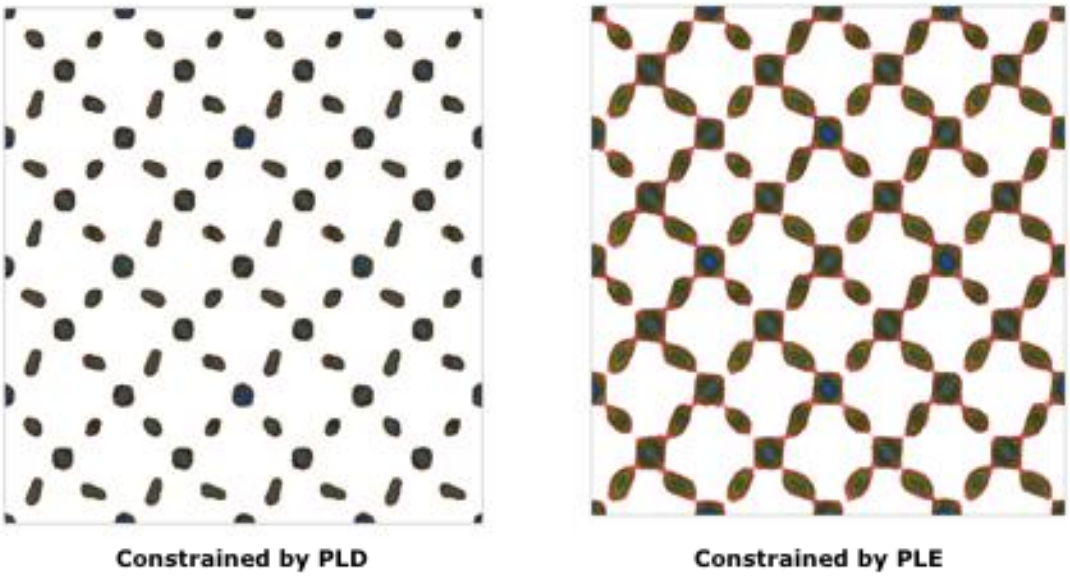


Figure 5.28 – Scheme showing dynamic connectivity for methane, using a probe radius of 2.126 Å. The dynamic connectivity for the RBM MD simulations is shown on the left, whilst the results for the FBM MD simulations are shown on the right. The pore was open for red = 1 ps (0.1%), green = 10 ps (1%) and blue = 100 ps (10%) of a 1000 ps production run.

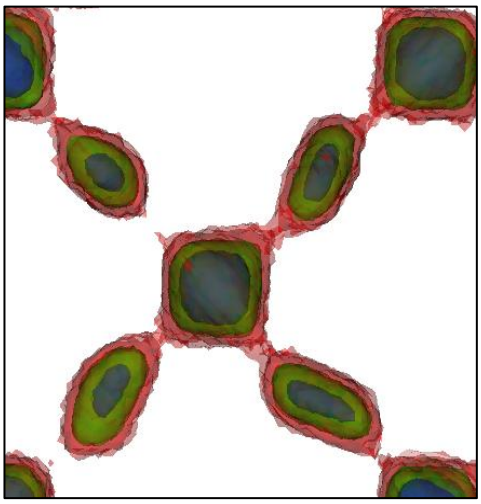


Figure 5.29 – Close up of the pore topology for the FBM MD simulation of methane, showing both connections and disconnections in the connectivity. The pore was open for red = 1 ps (0.1%), green = 10 ps (1%) and blue = 100 ps (10%) of a 1000 ps production run.

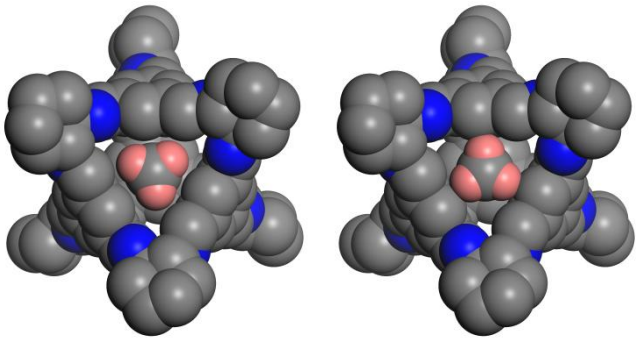


Figure 5.30 – Comparing the end-on to side profile for methane. By orientating the methane in the correct manner, it can fit within the window. All atoms were scaled to their CCDC radii and all hydrogens on the cage have been omitted for clarity.

The methane could diffuse through the window by moving one hydrogen atom through the window first and then reorientating so that the other three hydrogen atoms were aligned to the triangular cavity of the cage window before diffusing through. As the hydrogen atoms were able to re-orientate to fit through the window, the limiting diameter now becomes that of the central carbon atom. Figure 5.31 shows that if the radius of a single carbon (1.70 Å) was used to represent the methane, the pore connectivity was now fully connected. This helps to explain the experimental gas uptake observed for methane in **CC3**.

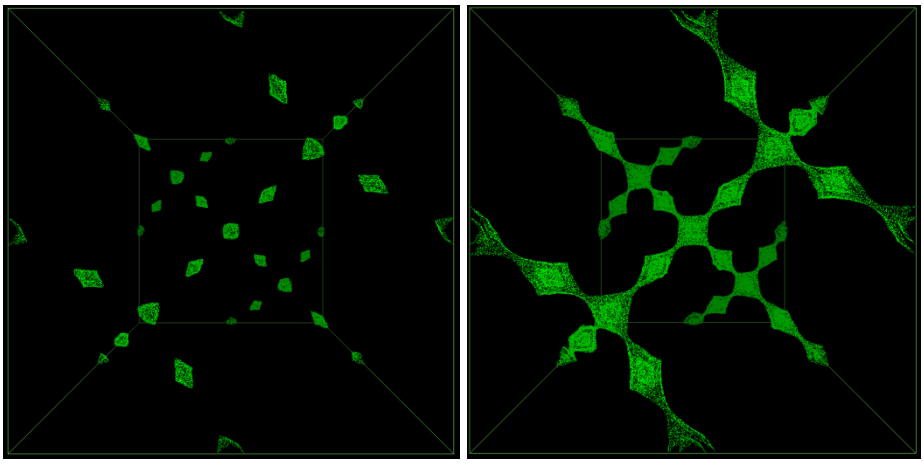


Figure 5.31 – Figure showing the pore topology for methane using the original spherical Van der Waal radius (2.126 Å) and then the Van der Waal radius when shape selectivity was considered (1.70 Å). This helps explain why diffusion of methane through **CC3** was observed in the rigid MD simulations.

During the RBM MD simulation, the methane occupies 95% of the cage cavities, Figure 5.32a. The trajectory of methane was shown in Figure 5.32b and shows how the gas twists in such a way to always be correctly aligned with the cage window, so that it can diffuse from cage, to interstitial site, to cage. Like carbon dioxide, it has to reorientate itself so that it fits through the window. Figure 5.32c shows this reorientation; note how one hydrogen points directly into the window, while the other three point towards the cyclohexane diamine vertices surrounding

the window. Like carbon dioxide, the diffusion must be either opportunistic or cooperative, and FBM MD simulations were carried out to provide further information about the system. The FBM MD simulations for methane in **CC3** show that 100% of the cages were occupied during the simulation; this was slightly more than in the RDM MD simulations (95%), and can be seen in Figure 5.32d. This suggests that the increased flexibility of the FBM helps to aid the gas diffusion somewhat.

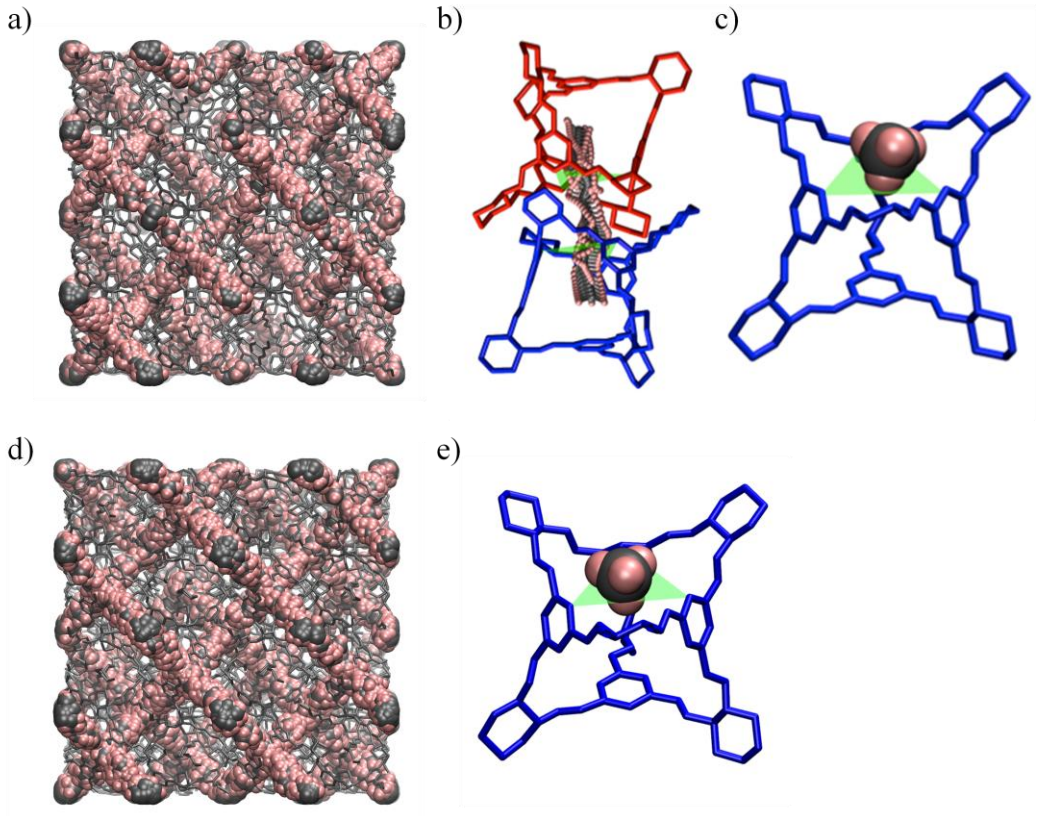


Figure 5.32 – This illustrates RDM MD simulations of, a) the position of CH_4 at every point during the simulation, b) the trajectory of a CH_4 molecule through the periphery of the cage window and, c) its orientation of the gas as it diffuses through the window of **CC3**. For the FBM MD simulations illustrations of d) the position of CH_4 at every point during the simulation, and e) its orientation of the gas as it diffuses through the window of **CC3**. All cages have been given a single colour, so that they stand out.

The shape exclusive nature of methane was seen in Figure 5.32c. The methane rotation and orientation can be illustrated by using the cartesian coordinates for the hydrogen atoms, obtained from the RBM MD simulation, as edges of a tetrahedron. This shows how the gas rotates so that one end of the tetrahedron fits through the window first, allowing the other corners of the tetrahedron to align with the gaps in the window, before passing through to the other side, Figure 5.33.

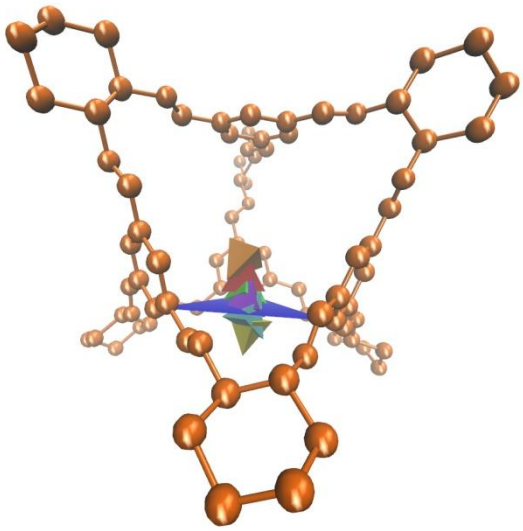


Figure 5.33 – Figure showing the re-orientation of a methane molecule through the **CC3** window. The gas has been drawn as a tetrahedron for simplicity. The gas was coloured to show its position at different points in time; this proceeds orange, red, green, purple, cyan and finally yellow.

The diffusion rates for the two simulations were again different, with the FBM MD simulation showing enhanced diffusivity. The diffusion coefficients were 9.6 x 10^{-9} m^2 s^{-1} and 5.39 x 10^{-9} m^2 s^{-1}, for the FBM and RBM simulations, respectively. The increased flexibility of the system allows the gas to diffuse more quickly; therefore the PLE may play a role in this. As more cages were occupied during the simulation, it seems to move at a faster rate than carbon dioxide; this may be because there are four possible orientations of the gas that can fit in the cage window, one for each hydrogen atom. Carbon dioxide by contrast has only two end-on orientations, and therefore diffuses at a slower rate. The only other factor that could alter the diffusion rates, is the minimum size of the gases, though this was in fact the same for both (1.70 Å). To examine whether cooperative diffusion was taking place, τ_0 was again used as a cut off for the PLE, Figure 5.35a. It was evident that as τ_0 was decreased, there was an effect on the PLE; when τ_0 was below 2 Å the PLE starts to shift to the right. To see whether this was a case of cooperative diffusion or just the gas taking advantage of the situation, the deviation from the crystallographic window diameter (3.62 Å) was again measured as a function of time and plotted on the same axis as τ_0 versus time, Figure 5.35b. As there was no significant deviation in the window diameter as methane passes through, it was clear that methane was diffusing through the system without directly affecting the cage window, hence exhibiting opportunistic diffusion. The shift in the PLE at small values of τ_0 seen in Figure 5.35a could account for the increased diffusivity observed in the FBM MD simulations.

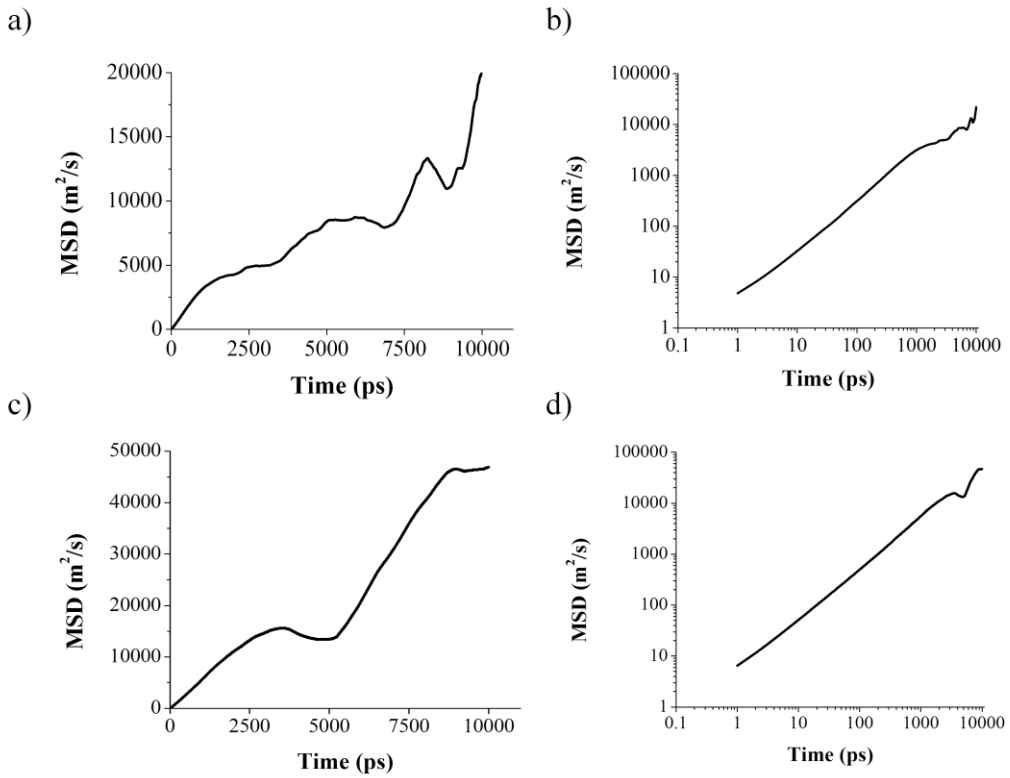


Figure 5.34 – The MSD for methane in **CC3** for a) the RBM simulation, b) the logarithmic plot of this to check normal diffusion had been achieved, c) the FBM MD simulation, and d) logarithmic plot for the FBM MD simulation.

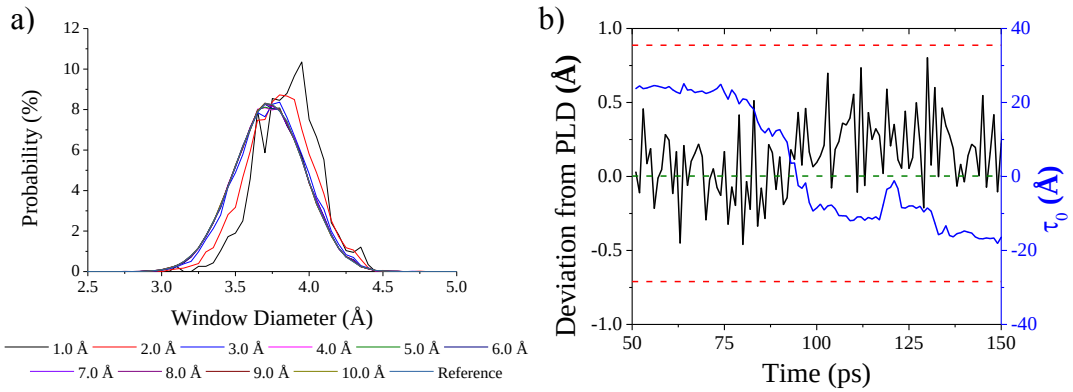


Figure 5.35 – Graph showing a) the PLE for a system containing methane for decreasing values of τ_0 compared to the PLE of the empty **CC3** reference system and b) the deviation of the window diameter from the original (3.62 Å) as a function of time (black) plotted against τ_0 versus time (blue). The dashed red lines show the maximum and minimum deviations, whilst the origin was marked with a dashed green line.

5.8.2. Sulfur hexafluoride

One molecule of sulfur hexafluoride was inserted into the centre of a cage within a 2 x 2 x 2 supercell of **CC3**. The size of the gas clearly limited the diffusion; it was far larger than the PLD of **CC3** with a spherical diameter of 6.05 Å. This was

much too large to fit through the cage window, and it was unsurprising that no experimental sorption isotherms could be calculated. In addition, no pore connectivity was seen when the static structure was examined. To see whether there was any dynamic connectivity, a probe radius of 3.03 Å was used as a screen, for both the RBM and FBM MD simulations, though no pore topologies were generated and therefore the 3D histograms could not be displayed. This suggests that the gas was too large to diffuse through the system and would therefore sit within a single cavity of the cage during the MD simulations. Figure 5.36 shows how sulfur hexafluoride does not fit within the cage window, suggesting that the window would have to swell a great deal if the gas was to diffuse.

Figure 5.36 – Comparing the end-on to side profile for sulfur hexafluoride (in purple). No orientation of the gas allows it to fit in the cage window. All atoms were scaled to their CCDC radii and all hydrogens on the cage have been omitted for clarity.

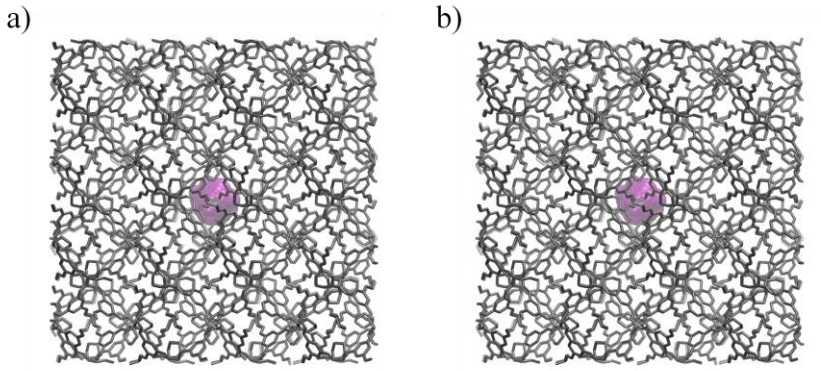


Figure 5.37 – Illustration showing the superimposition of sulfur hexafluoride during a) the RBM MD simulation and b) the FBM MD simulation. It was evident that the gas does not diffuse through **CC3**.

However since methane was able to diffuse successfully through **CC3** during the RBM MD simulation, it was therefore important to confirm that sulfur hexafluoride cannot diffuse in flexible simulations. Both the RDM and FBM MD simulations showed similar results; the gas just oscillates within the centre of one cage, Figure 5.37. This confirms that sulfur hexafluoride is not able to diffuse through **CC3**. Any calculated diffusion coefficients were therefore meaningless, as the normal diffusion rate could not be reached.

5.9. MD simulations of spherical gases

Two spherical, monatomic gases were simulated within **CC3**; krypton and xenon. Although similar in size, initial screening using methods proposed by Sikora *et al.*[18] suggested that **CC3** might be ideal for separation of these noble gases. This is interesting because noble gas separation is significant on an industrial scale, and these gases are difficult to separate due to their lack chemical reactivity, low conductivity, and low melting and boiling points.[5] Experimental isotherms, however, seem to contradict this simple conclusion, showing uptake of both gases.

5.9.1. Krypton

For krypton, experimental sorption isotherms at a pressure of 1 bar and a temperature of 263 K, Chapter 1, show an uptake of around 1.8 mmol g^{-1}. This is interesting as the size of krypton is almost identical to the PLD in **CC3**. This means that the topology of the static picture was almost connected, Figure 5.1e. The dynamic connectivity for krypton in both the RBM and the FBM MD simulations were screened using a probe radius of 1.818 Å. Initial viewing of the topologies generated for both the RBM And FBM MD simulations would suggest full connectivity, as seen in Figure 5.38, however closer inspection of the RBM pore topology shows that this was in fact disconnected, Figure 5.39a. The reason for this was that the probe radius used was slightly larger than the PLD; the probe diameter was 3.64 Å, versus 3.62 Å for the PLD. As krypton was modelled using a potential, there was scope for the effective radius of the atom to shrink slightly in the simulation. As the difference between the probe diameter and the PLD was only 0.44%, this falls within the error of the calculations and diffusion of krypton through the RBM MD simulation could be possible. This means that krypton was a tight fit in the cage window, Figure 5.39b. The actual size of the pore channels seen for both the RBM and FBM simulations differed greatly; the channels seen for the FBM results are far more pronounced. In these, the blue isosurface shows that the channels are open for longer periods of time, with the channels between the cage cavity and interstitial sites being almost as large as the cage cavity show in the RBM results.

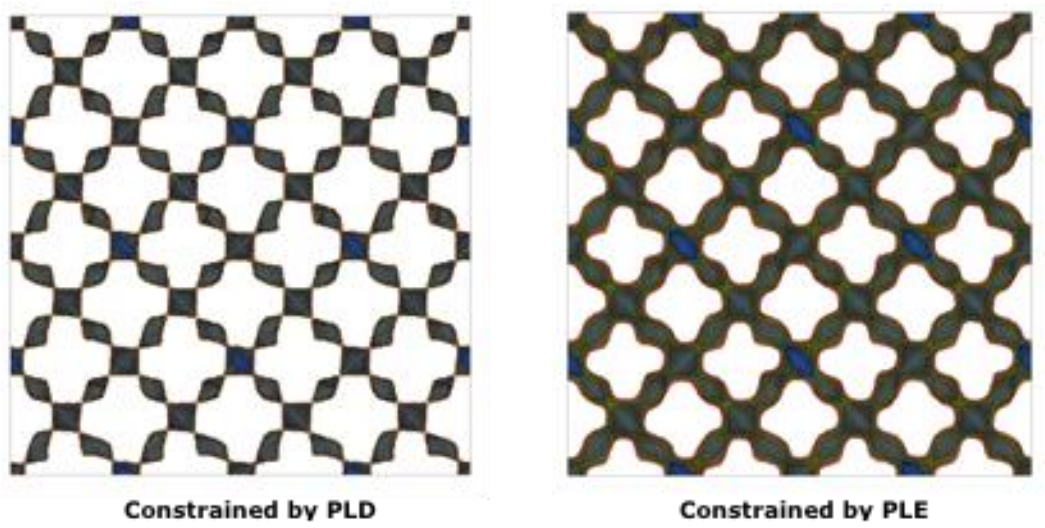


Figure 5.38 – Scheme showing dynamic connectivity for krypton, using a probe radius of 1.818 Å. The dynamic connectivity for the RBM MD simulations is shown on the left, whilst the results for the FBM MD simulations are shown on the right. The pore was open for red = 1 ps (0.1%), green = 10 ps (1%) and blue = 100 ps (10%) of a 1000 ps production run.

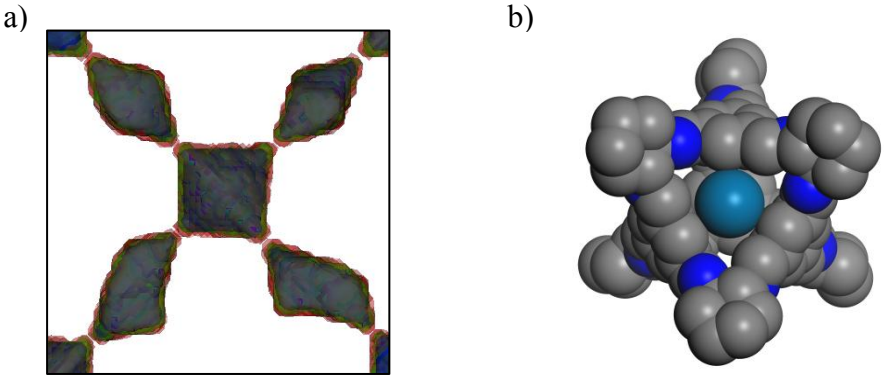


Figure 5.39 – a) Close up of the pore topology for the RBM MD simulation of krypton, showing no connectivity due to the size of the probe diameter exceeding the PLD. The pore was open for red = 1ps, green = 10 ps and blue = 100 ps of a 1000 ps production run. b) Figure showing krypton in the centre of the cage. All atoms were scaled to their CCDC radii and all hydrogens on the cage have been omitted for clarity.

To investigate the diffusivity of krypton, it was inserted into the centre of a cage within a 2 x 2 x 2 supercell of **CC3**, and both RBM and FBM MD were simulated. The RBM MD simulations show that krypton can diffuse through **CC3**, occupying 78% of the cage cavities, Figure 5.40a. As the krypton was not modelled as a hard sphere and was only slightly different in size to the PLD, this was not surprising. The trajectory of krypton through the pair of cages, Figure 5.40b, shows that it dissects the window in the centre, suggesting that the fit was either a very tight, or this pathway maximises the van der Waal interactions. It was important to compare these results to the FBM MD simulations before assuming this was another example of passive diffusion.

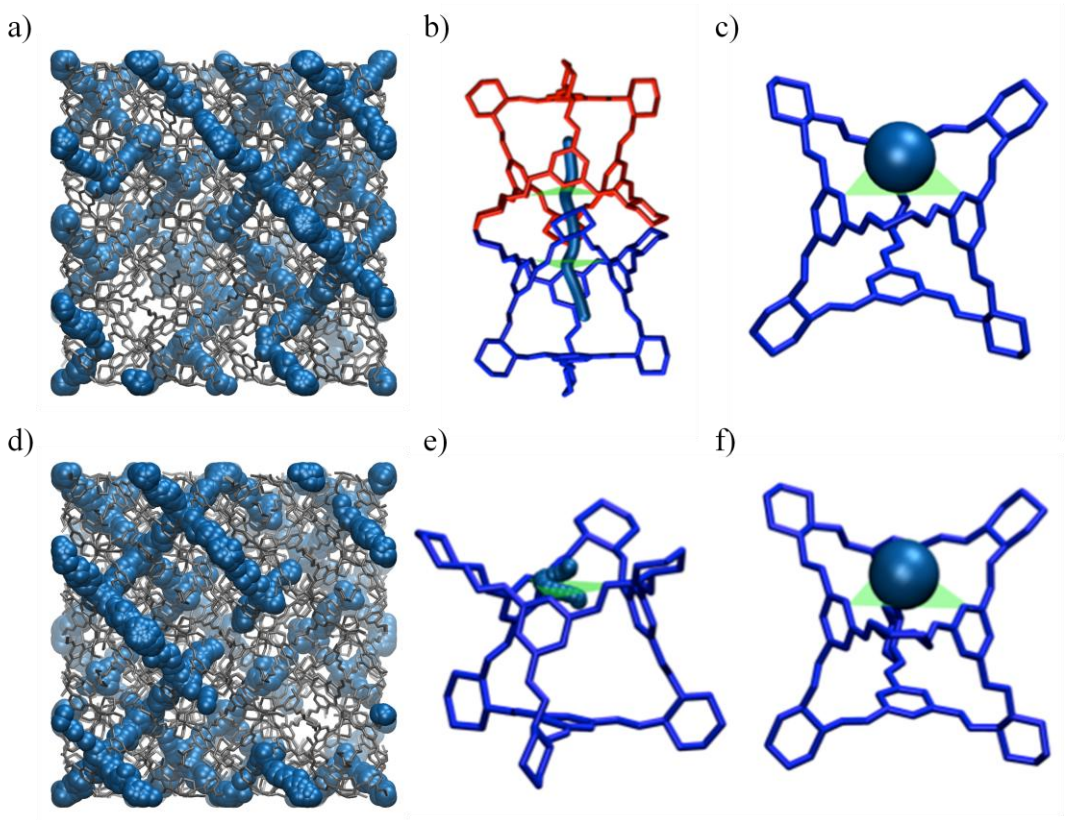


Figure 5.40 – This shows illustrations for RDM MD simulations of the a) position of krypton at every point during the simulation, b) the trajectory of a krypton molecule through the periphery of the cage window and c) its orientation of the gas as it diffuses through the window of **CC3**. For the FBM MD simulations illustrations of d) the position of krypton at every point during the simulation, e) the trajectory of a krypton molecule through the periphery of the cage window and f) its orientation of the gas as it diffuses through the window of **CC3**. All cages have been given a single colour, so that they stand out.

The FBM MD simulation shows similar results to the RBM counterpart. However, the gas occupies just under 10% more of the cages, at 86%, Figure 5.40d; this was likely to be a result of the increased flexibility. The diffusion profile of the gas, Figure 5.40e, shows that the diffusion mechanism has been altered. In the RBM system krypton passes directly through the centre of the cage window, whereas in the FBM simulations it was able to move through the periphery of the window. This was shown in Figure 5.40f. It was interesting that the diffusivity of the two different methods was similar; 2.09 x 10^{-9} m^2 s^{-1} for the RBM simulations and 2.40 x 10^{-9} m^2 s^{-1} for the FBM simulations. This suggests that the overall motion of the gas was not affected by the host structure.

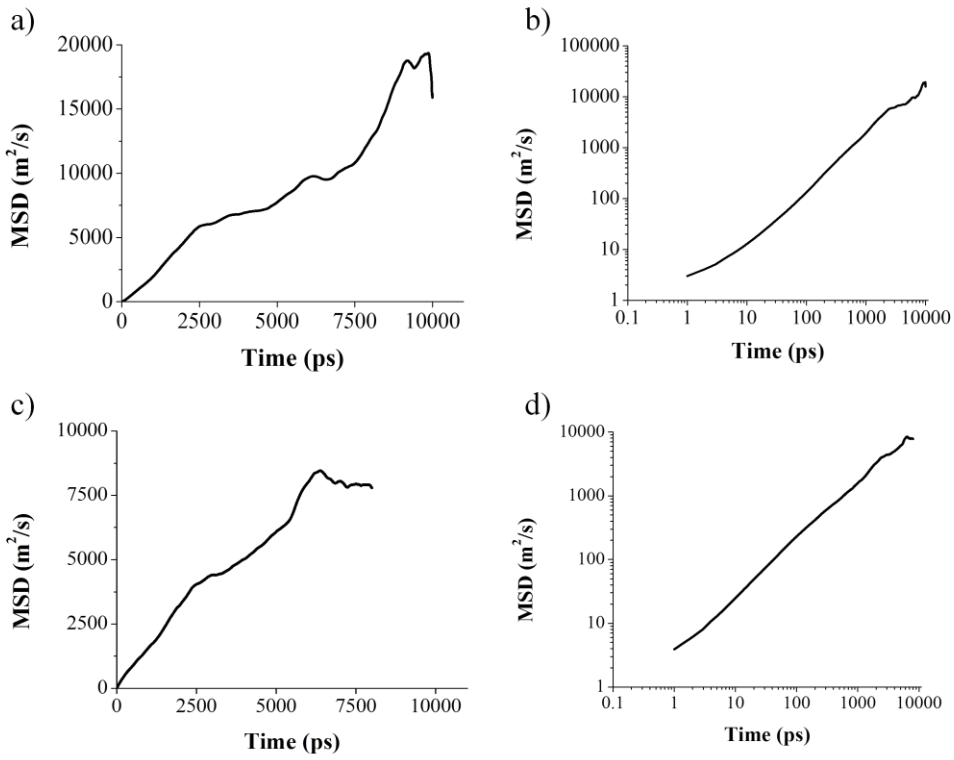


Figure 5.41 – The MSD for krypton in **CC3** for a) the RBM simulation, b) the logarithmic plot of this to check normal diffusion had been achieved, c) the FBM MD simulation, and d) logarithmic plot for the FBM MD simulation.

When considering the effect the gas has on the window, it was clear that the PLE was being directly affected as the gas moves into a closer proximity, Figure 5.42a. When τ_0 = 3.0 Å there was a slight change in the distribution and this shift was increased as τ_0 tends towards 1.0 Å. It seems that when τ_0 was at this minimum point, the range was also shifted, so that the minimum window diameter has to be open slightly wider (3.20 Å). Like the reference data, the maximum value for the window diameter does not exceed 5.0 Å; only the distribution of the PLE was altered. Since krypton moves through both the RBM and FBM MD simulations, it was difficult to define this as cooperative diffusion, however it was prudent to check how the window behaves as the gas moves through it. The deviation from the static window diameter (3.62 Å) was plotted as a function of time, and drawn on the same axis as τ_0 versus time, Figure 5.42b. Even though two events of the gas moving though the cage window were shown, the window deviation was not affected. This therefore suggests opportunistic diffusion. Like methane, shifts in the PLE at small values of τ_0, seen in Figure 5.42a, must be down to the gas taking advantage of the window being open, opposed to forcing its way through. This would also account for the increased diffusivity observed in the FBM MD simulations.

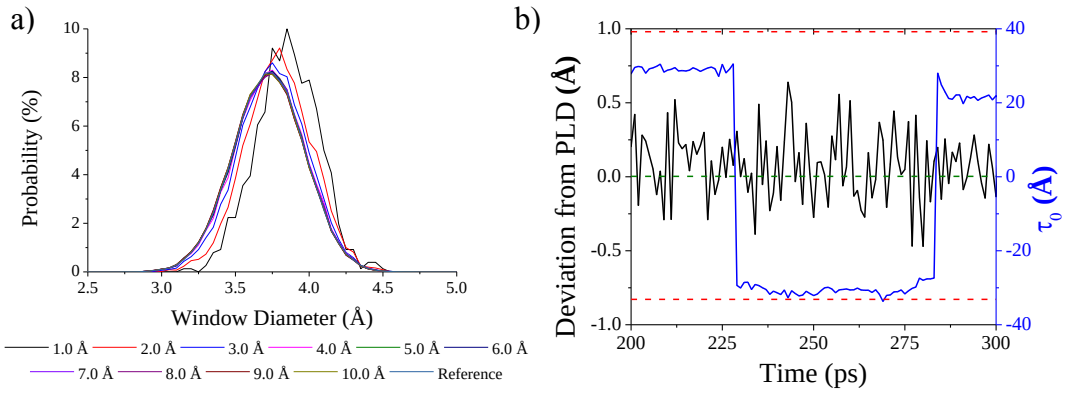


Figure 5.42 – Graph showing a) the PLE for a system containing krypton for decreasing values of τ_0 compared to the PLE of the empty **CC3** reference system and b) the deviation of the window diameter from the original (3.62 Å) as a function of time (black) plotted against τ_0 versus time (blue). The dashed red lines show the maximum and minimum deviations, whilst the origin was marked with a dashed green line.

5.9.2. Xenon

The initial static pore topology (Figure 5.1f) suggested that xenon would *not* be able be able to diffuse through the cage system, as only disconnected voids were present, however experimental sorption isotherms show that, at a pressure of 1 bar and a temperature of 263 K, there is a sharp uptake of around 2.6 mmol g^{-1}. It was important to see whether dynamic connectivity could explain this. Isosurfaces were generated for both MD simulations using a probe radius of 2.05 Å; this was much larger than the PLD of the system, and therefore the pore topology for the RBM MD simulation shows disconnected cavities – one for the cage cavity, the other for the interstitial sites. The FBM MD simulations, though, show a fully connected 3D diamondoid pore topology; this suggests the flexibility plays a key role in xenon diffusion through **CC3**. The topologies for both simulations were shown in Figure 5.43.

It was also possible to compare two gases, with the view to assess the residence time of the pore topologies, and whether this would have an impact on the kinetics of the gases. We expected that the longer the topology was connected, the faster the diffusion of the gas through the system. To illustrate this concept, the two noble gases, krypton and xenon, were compared. As the pore topology of xenon was disconnected for the RBM MD simulations, only the results for the FBM MD simulations have been shown; Figure 5.44. It was clear that krypton displays a more connected 3D pore topology for a greater extent of the production run; this was unsurprising due its reduced size when compared to xenon. This would suggest that krypton would diffuse through the system at a faster rate than the

xenon. As xenon is spherical in nature, this cannot be ascribed to a shape effect, as for methane, and this was confirmed when a xenon atom was put in the centre of the cage window, Figure 5.45.

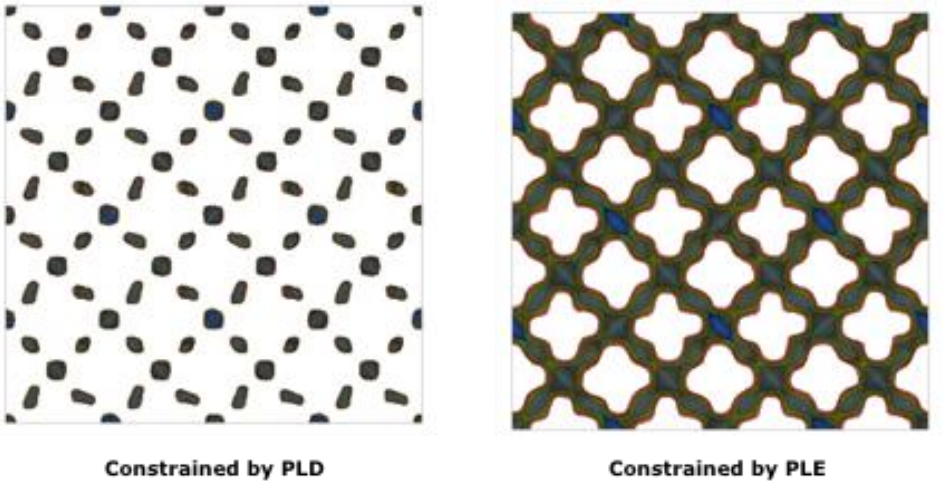


Figure 5.43 – Scheme showing dynamic connectivity for xenon, using a probe radius of 2.05 Å. The dynamic connectivity for the RBM MD simulations is shown on the left, whilst the results for the FBM MD simulations are shown on the right. The pore was open for red = 1 ps (0.1%), green = 10 ps (1%) and blue = 100 ps (10%) of a 1000 ps production run.

The RBM MD simulation shows that xenon would not be able to diffuse through the system. Superimposition of the trajectory of the gas, Figure 5.46a, shows that the gas remains within the centre of one cage throughout the simulation; this correlates to a cage occupancy of just 1.56%. Figure 5.46b shows the trajectory of the xenon as it attempts to leave the cage, but reaches the cage window, was too large to transverse it, so therefore returns back to the centre of the cage. This implies that the RBM MD simulation was incapable of describing the diffusion of xenon through **CC3**, therefore suggesting the increased flexibility in the FBM was vital for describing the diffusivity. The diffusion coefficient for the RBM MD simulations could not be determined, as the normal diffusion regime was never achieved. The superimposition of the xenon position during the FBM MD simulation, Figure 5.46c, to show that xenon now diffuses throughout the simulation.

The diffusion trajectory for xenon, Figure 5.46d, shows diffusion through the centre of the cage window and it occupies 10.94% of the cage cavities during the full simulation. Xenon always passes through the centre of the cage window, as shown in Figure 5.46e. This shows that when the flexibility of the system was considered, xenon does diffuse through **CC3**. The diffusion coefficient for xenon in the FBM MD simulation was calculated to be 1.83 x 10^{-10} m^2 s^{-1}. This was slower than all of the other gases simulated so far that diffuse through **CC3**, and this

was a result of the size of the gas being very close to the PLD. As xenon was the largest gas to diffuse through **CC3**, it was the most likely gas to show some kind of cooperative diffusion.

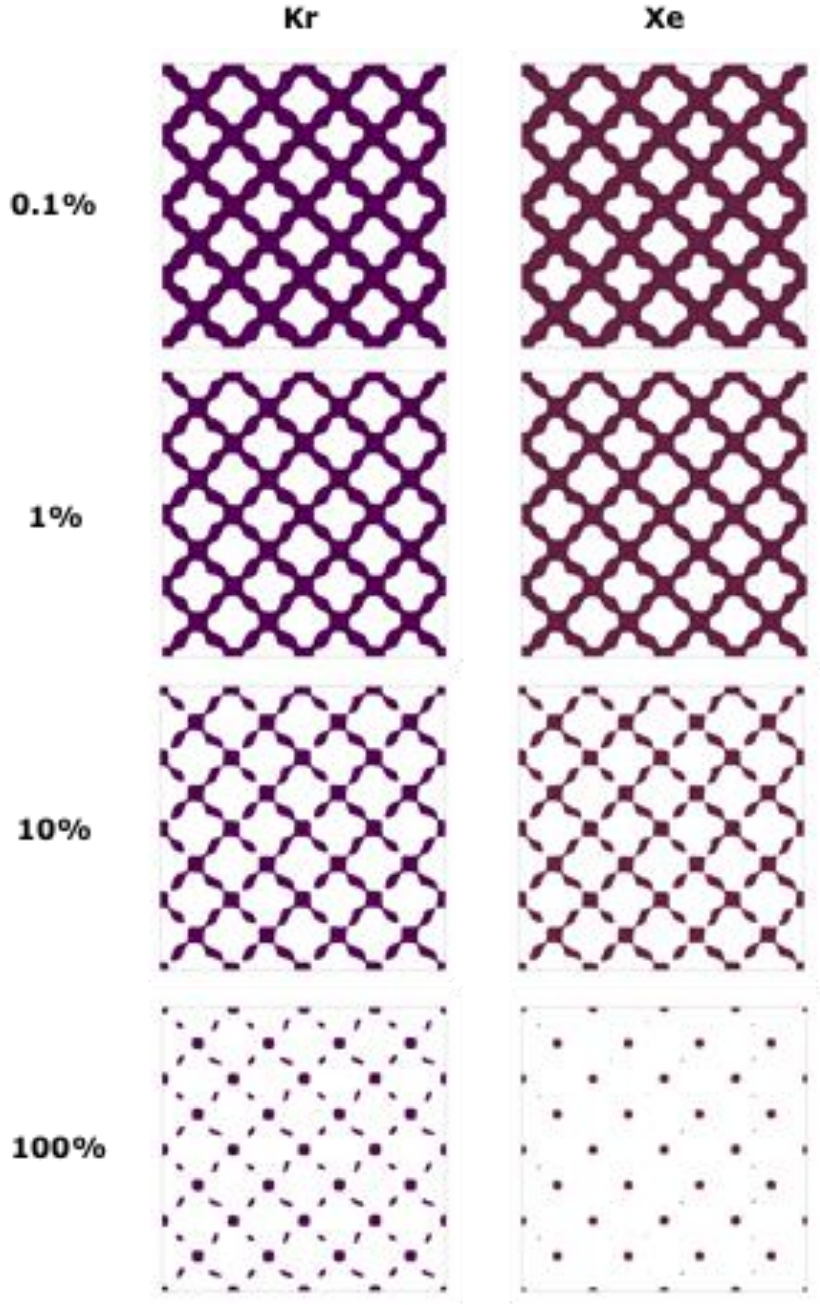


Figure 5.44 – Scheme showing the pore topology for krypton and xenon using the FBM MD simulation data, using a probe radius of 1.818 Å and 2.05 Å respectively. The percentage of time the topology was open for over a 1000 ps simulation was shown.

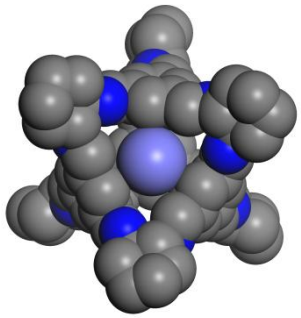


Figure 5.45 – Figure showing xenon placed in the centre of the cage window. All atoms were scaled to their CCDC radii and all hydrogens on the cage have been omitted for clarity.

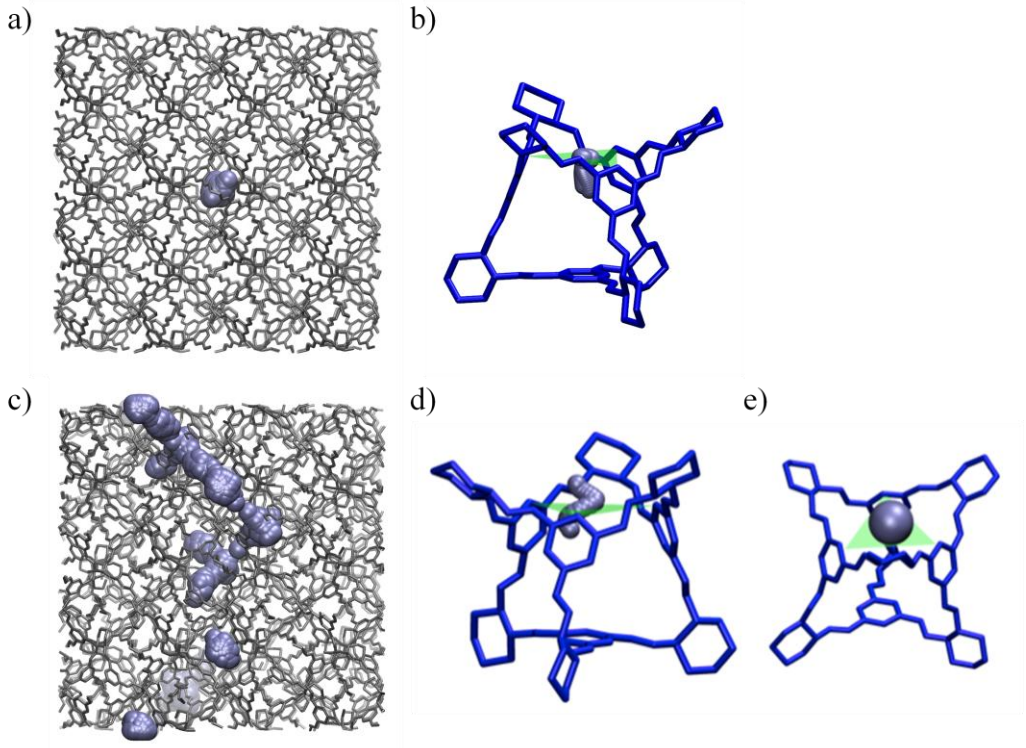


Figure 5.46 – This shows illustrations for RDM MD simulations of the a) position of xenon at every point during the simulation, b) the trajectory of a xenon molecule through the periphery of the cage window. For the FBM MD simulations illustrations of c) the position of xenon at every point during the simulation, d) the trajectory of a xenon molecule through the periphery of the cage window and, e) its orientation of the gas as it diffuses through the window of **CC3**. All cages have been given a single colour, so that they stand out.

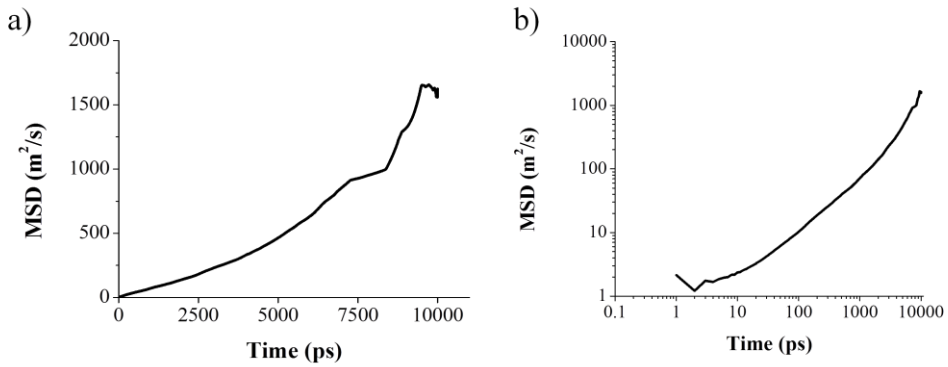


Figure 5.47 – The MSD for xenon in **CC3** for a) the FBM simulation and b) the logarithmic plot of this to check normal diffusion had been achieved.

To see whether cooperative diffusion was taking place, τ_0 was used as a screening criterion for the PLE and the results compared, Figure 5.48a. It was evident that there was a large change in the PLE when τ_0 was reduced; when τ_0 = 1.0 Å, the PLE shifted to the right, suggesting that the gas had an effect on the distribution. The jagged nature of this plot was a consequence of the quality of data being restricted such that a smooth curve could not be successfully drawn; this still shows, though, that the PLE has a different distribution to that expected. If this information was combined with the fact that xenon does not diffuse through the RDM MD simulations, it suggests that cooperative diffusion could be taking place. Figure 5.48b, though, disproves this theory: like all the other gases, the window diameter was unaffected as the gas travels through the cage window. This means that the cages were not influenced by the guests inside them, and that xenon exhibits opportunistic diffusion; it only moves through when the window was large enough but does not force an abnormal change in the window diameter to facilitate this movement.

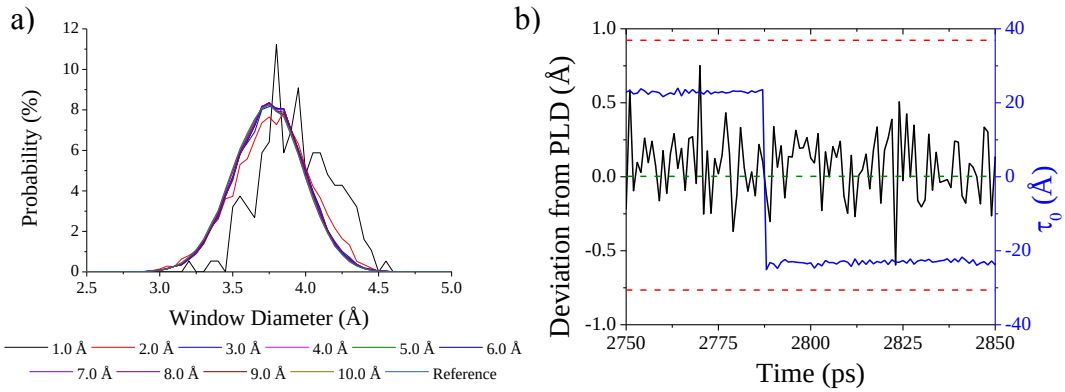


Figure 5.48 – Graph showing a) the PLE for a system containing xenon for decreasing values of τ_0 compared to the PLE of the empty **CC3** reference system and b) the deviation of the window diameter from the original (3.62 Å) as a function of time (black) plotted against τ_0 versus time (blue). The dashed red lines show the maximum and minimum deviations, whilst the origin was marked with a dashed green line.

5.10. Thermodynamic calculations for the noble gases

The kinetic diffusivities for the noble gases suggest that krypton has the ability to diffuse through **CC3** at faster rate than xenon. That said, experimental results show that, at one bar, there is a higher uptake of xenon than krypton. To try and understand this, single component sorption isotherms were calculated for both gases. Subsequently, competitive dual-component sorption were also simulated. The aim here was to see whether there is an energetic preference for the uptake of xenon over krypton, even though it was diffusing at a slower rate.

Considering the single component sorption isotherms, the initial slopes for both the krypton and xenon simulated isotherms, Figure 5.49a and 5.49b respectively, show that after 100 kPa **CC3** was calculated to have an uptake of 3.29 kryptons per cage and 3.00 xenons per cage. For the same pressure experimentally, **CC3** shows an uptake of only 2.22 kryptons per cage and 3.16 xenons per cage. This suggests that the xenon isotherm was showing a close agreement with the saturation uptake observed in **CC3**, while it was clear that simulations were overestimating krypton uptake.

Looking at the profile of the isotherms, it was clear that for both krypton and xenon the uptake was initially being overestimated, as saturation was being reached too quickly. One reason for this discrepancy was that a perfect and infinite crystal was used to calculate the simulated sorption, whereas the experimental sorptions rely heavily on the crystallinity of the sample. Historically it has been shown that **CC3** can be amorphised, so much so that great care has to be taken to prepare a sample of the system otherwise crystallinity can be lost.

This could help explain the differences between the calculations can be expected.[36] In addition, the system was considered rigid during the GCMC calculations and previous kinetic studies have shown the necessity of system flexibility, especially for the noble gases. Finally Li *et al.*[37] have recently shown how difficult it was to model isotherms for microporous organic molecular crystals. Here accurate *ab initio* potentials were used as reference data for a custom fitted FF and this still showed an overestimation in the sorption isotherms. However, for **CC3** what does look reasonable was the total uptake of xenon, as the simulated and experimental sorption isotherms seem to agree that a saturation uptake of around 3 xenons per cage was possible in **CC3**. This was unsurprising, as geometrically this would be expected; one in the centre of the cage and then half in each of the four cage windows. With krypton, it looks as though the experimental sorption isotherm may not yet have reached equilibration, Figure 5.49c; therefore, this could potentially reach the suggested simulated uptake of around 4 kryptons per cage.

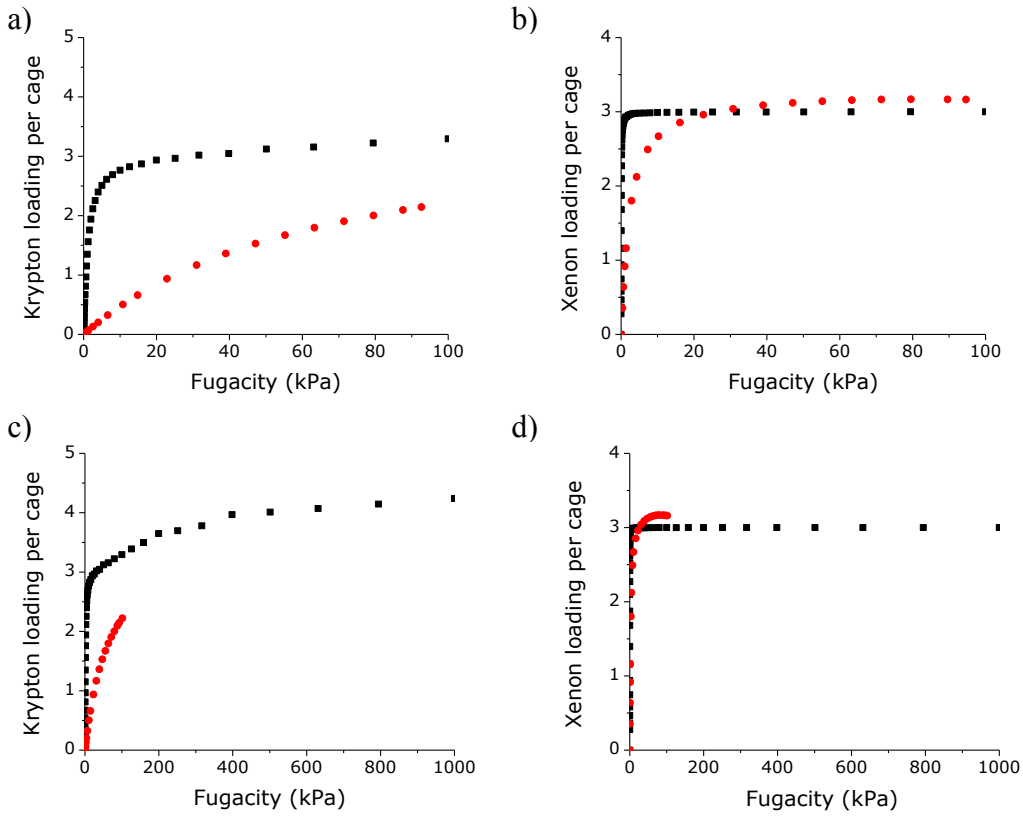


Figure 5.49 – Figure showing the simulated (black) and experimental (red) sorption data for a) krypton over a pressure range of 0.1 to 100 kPa, b) xenon over a pressure range of 0.1 to 100 kPa, c) krypton over a pressure range of 0.1 to 1000 kPa and d) xenon over a pressure range of 0.1 to 1000 kPa.

An inspection of the density field maps allows the preferred sorption sites for the gases to be determined. These display where the gases were adsorbed within the system as a function of a normalised frequency; where 1.0 was the field

maximum, at a pressure of 10 kPa. Figure 5.50 displays the density field maps for both krypton and xenon. It was clear that there were specific sites for xenon, whereas there were no specific sorption sites for krypton, instead it follows the pore topology. Xenon has a higher sorption energy then krypton, Figure 5.51. What is interesting is that xenon seems to have two sorption sites, both of which have a more favourable energy than krypton. By looking at the individual structures, it was possible to deconvolute the two different sorption sites. The higher energy sorption site, with a sorption energy of ~43 kJ mol^{-1}, was the cage interior, whereas the slightly lower adsorption site, with a sorption energy of ~40 kJ mol^{-1}, was when the gas was found in the interstitial site. The calculated isosteric heats at 273 K for the gases were 31.05 kJ mol^{-1} and 41.86 kJ mol^{-1} for krypton and xenon respectively. These were higher than the experimental isosteric heats values; krypton = 24.4 kJ mol^{-1}, xenon = 29.6 kJ mol^{-1}. This would agree with earlier insights that the intermolecular interactions may be being overestimated.

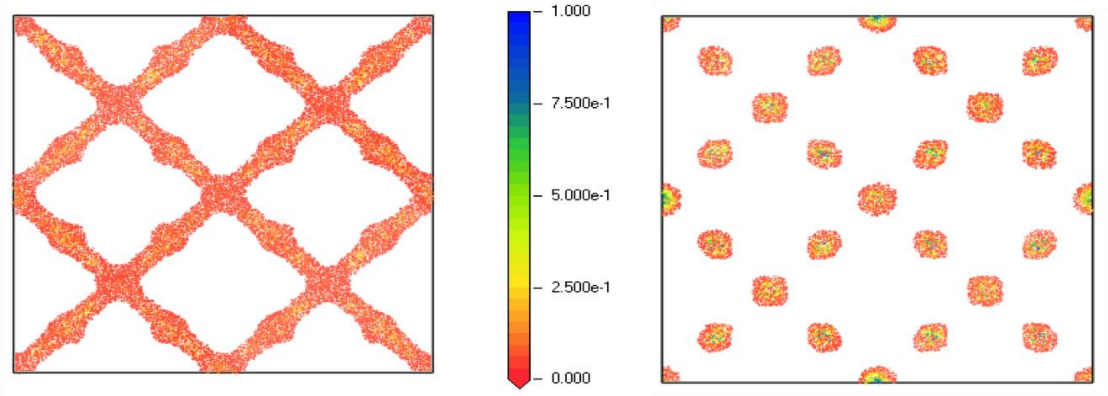


Figure 5.50 – Density field maps for krypton (left) and xenon (right). These display where the gases were adsorbed within the system as a function of a normalised frequency; where 1.0 was the field maximum, at a pressure of 10 kPa. Krypton shows no specific adsorption sites, whereas xenon shows distinct sites; these were located in the cage cavity and interstitial site.

A competitive dual-component sorption calculation was also studied, to see whether the **CC3** system shows preference to one noble gas over the other. This was carried out with an equal ratio of both gases, at a fixed pressure of 10 kPa; all other conditions matched the single component sorption isotherm calculations. The results show that there was a high preference for xenon uptake. The average loading of the cell was 22.55 xenon to just 1.28 krypton; that was 2.82 xenon per age to 0.16 krypton per cage. This gives a selectivity ratio of around 17.63:1, xenon:krypton. By looking at the results in more detail, it was evident that although both noble gases compete at the two cage sites (cage cavity and interstitial site), the cage cavity strongly favours xenon. This was unsurprising, given the sorption energies shown in Figure 5.51.

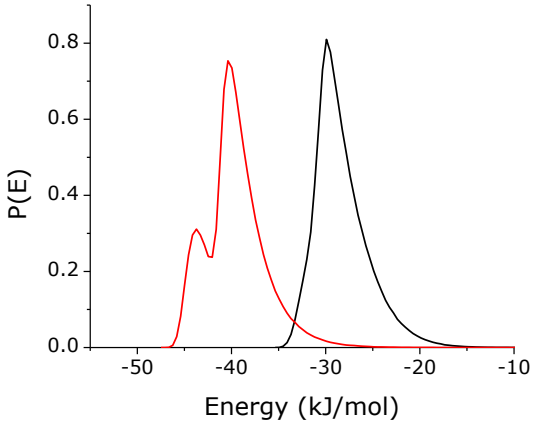


Figure 5.51 – Graph showing the different sorption sites for krypton (black) and xenon (red). Xenon shows two distinct sorption sites, one at ~43 kJ mol^{-1} and one at ~40 kJ mol^{-1}. These correspond to the cage cavity and interstitial site respectively.

Overall although krypton has a higher diffusivity, uptake of xenon was shown to be preferential, due to higher sorption energies, higher isosteric heats as well as specific sites within the system for sorption. The experimental isotherms are measured, of course, at equilibrium.

5.11. Comparison of the kinetic results

To see if there were any trends in the kinetic results, the diffusion coefficients for the gases were examined, and their values for both the RBM and FBM compared; this is tabulated in Table 5.2. The ratio of the RBM to the FBM simulations was also included, so that the effect of flexibility on the rate of diffusion could be quantified. This helped quantify what effect, if any, the increased flexibility of the simulations was having on the rate of diffusion. It was expected that this difference would become more prominent the larger the gas molecule, as the PLD would become a limiting factor.

When considering hydrogen, it was clear that the increased flexibility in the FBM MD simulation had little impact on the rate of diffusion and this suggests that hydrogen was too small to be effected by the PLD and therefore the PLD was not a limiting factor. This would imply that the FBM MD simulation was not required to describe hydrogen and that treating the cages as rigid bodies was an accurate assumption. As hydrogen is the smallest of the linear gases this was unsurprising. Nitrogen was the second smallest linear gas, but the small increase in size does seem to have an impact on the diffusion rate. Unlike hydrogen, which had almost identical diffusion rates for both MD simulations, the ratio for nitrogen shows that there was a difference. The increased flexibility of the FBM MD simulation improved the rate of diffusion by almost 20%. This would suggest that the PLD

has started to play a role in the diffusion of the gas and therefore the FBM MD simulations were required.

Gas	Min Gas Size (Å)	Atomic weight (u)	Is re-orientation required?	RBM diffusion coefficient (m^2 s^{-1})	FBM diffusion coefficient (m^2 s^{-1})	Ratio
H_2	2.18	2.02	No	$5.44x10^{-8}$	$5.64x10^{-8}$	1.04
N^2	3.10	28.01	No	$4.98x10^{-9}$	$5.94x10^{-9}$	1.19
CO_2	3.40	44.01	Yes	$9.72x10^{-10}$	$2.60x10^{-9}$	2.67
CH_4	3.40	16.04	Yes	$5.39x10^{-9}$	$9.60x10^{-9}$	1.78
Kr	3.64	83.80	No	$2.09x10^{-9}$	$2.40x10^{-9}$	1.15
Xe	4.10	131.29	No	N/A	$1.83x10^{-10}$	N/A

Table 5.2 – Table comparing the diffusion coefficients from both the RBM and FBM MD simulations for the different gases.

When the flexibility of the simulation was increased for the largest linear gas, carbon dioxide, the greatest change in diffusion rate was observed. The ratio of the diffusion coefficient for the RBM and FBM was 2.67, demonstrating how vital the enhanced flexibility in the system was. This was likely to be because carbon dioxide seemed to diffuse through the cage window in an end-on orientation (Figure 5.23), and this re-orientation may have slowed the diffusion. Introducing flexibility in the cage allows the size of the window to increase and therefore the gas was able to traverse the cage window more easily; this though was still in an end-on manner.

When comparing the linear gases, it was evident that there was a direct link between the size of gas and dependence on system flexibility; as the size of gas was increased, the system flexibility becomes more important. The explicit diffusion rate of the gases was also dependent on the size of the gas; hydrogen was the faster, then nitrogen, then carbon dioxide.

As methane is tetrahedral in nature, initial results would indicate that it adopts a specific orientation to pass through the cage window; one of the hydrogens is eclipsed with the carbon atom, while the other three hydrogens point into the free space within the cage window (Figure 5.30). When shape exclusivity was taken into consideration, the same probe radius was used to depict both carbon dioxide and methane. Therefore, it was crucial to compare the two gases and how their diffusion rates were altered as the flexibility of the system was increased. It was clear that for both the FBM MD simulations displayed an enhanced diffusion rate; the ratio for this increase though was different – the diffusion rate of carbon dioxide showed a greater dependence on the flexibility of the system. This could

be because carbon dioxide was limited to adopting an end-on profile as it diffused through the cage window; this gives two possible orientations. The tetrahedral symmetry of methane allows the adoption of four different orientations that will fit in the cage window, and this increases the likelihood of the gas being in the correct orientation as the cage window was approached; this could help explain the increased diffusion rate. This would suggest some type of shape preference.

The relative diffusion rate of the spherical gases, krypton and xenon, was dependent on their size and weight. Considering krypton first, this had a similar diffusion rate for both the RBM and FBM MD simulations; apart from hydrogen it was the gas least affected by the increased flexibility in the system. This is interesting, as initial calculations suggest that the diffusion of krypton through the system may be hindered by the PLD in the RBM simulations – this though seemed not to be the case. The fact that the diffusion rate was unaffected by the enhanced flexibility was therefore unsurprising, as it was spherical in nature so if it can diffuse through the RBM MD simulation, there was not much to be gained by allowing the PLD to fluctuate in size. Kryptons diffusion coefficient was also very similar to that seen for the FBM MD simulations of carbon dioxide.

Xenon did not diffuse through the RBM MD simulation, and this shows how important the flexibility was when the size of the gas is similar to the PLD. Recent work by Sikora *et al.*[18] would suggest by comparison that **CC3** would not allow xenon to diffuse through the pore network, but here we have shown that this assumption was invalid for these cage systems; this was validated by experimental sorption measurements of xenon in **CC3**, which saw uptake of the gas into the system.

The overall trend for the explicit rates of diffusion for the different gases, as seen in the FBM MD simulations, is:

Hydrogen > Methane > Nitrogen > Krypton > Carbon dioxide > Xenon

Instead of being defined solely by their size or shape, it more follows the general trend of their atomic weight; the lightest gas travels the fastest, the heaviest the slowest. This trend though was not perfect, as carbon dioxide was slightly slower than krypton, even though it has a lower atomic weight. Though would suggest that a combination of other factors, such as size and shape, have an impact on the relative rates of diffusion and that no single criteria determines the diffusion trend seen in **CC3**. By plotting the size of the gas against their diffusion coefficient, Figure 5.52, it was possible to cluster the gases into three groups.

The first group is when the gas is much smaller than the PLD; the only gas that fits into this group is hydrogen. The second group consists of all the gases that are just smaller or similar to the PLD; the gases that fit into this group include nitrogen, carbon dioxide, methane and krypton. The final group included any gas much larger than the PLD; the only gas in this group is xenon.

Group 1 has a must higher diffusion coefficient and this is because the dynamic motion of the cages does not have an affect on the gas involved; this can be rationalized by the small size of the gas in comparison to the PLD.

Group 3 has the lowest diffusion rate. This is because the gas, xenon, is dependent on the dynamic motion of the cages. If the cages do not 'breathe' the gas does not diffuse.

This leaves group 2. This is a cluster of gases, all with similar sizes. The differences in their diffusion rates can be explained by both differently contributions; namely their shape and size. The diffusion rates place these group 2 gases in the following order:

Methane > Nitrogen > Krypton > Carbon dioxide

If just the size is take into account, assuming that methane acts as a single carbon (the rationale for this was explained previously in Section 5.8.1), the following trend would be expected:

Methane = Carbon dioxide > Nitrogen > Krypton

Therefore, the disparity between these values could be due to the affect the shape of the gases has on the diffusion rate. First, as krypton is spherical it has a uniform shape and therefore the diffusion cannot be limited by different orientations of the gas. Next, nitrogen gas can fit in the cage window with a side-on orientation (Figure 5.23) showing that the shape is unlikely to affect the diffusion rate of the gas. This leaves both methane and carbon dioxide; it has been shown earlier that these must re-orientate to diffuse through the window. The trend in the gases for group 2 would suggest that this re-orientation could have a profound effect on the overall diffusion coefficients. Although methane still remains the fastest gas, carbon dioxide is the slowest out of the four gases, even though it should be similar to methane. However, when the shapes of the gases were taken into account, the reason for this becomes clear. As methane is a tetrahedral molecule this means that it has four different orientations that would fit in the cage window; carbon dioxide must have an end-on orientation to fit and

this only has two permutations. This means that methane is more likely to be in the correct configuration therefore increasing the probability of it diffusing from cage to cage; this would increase would account for the relatively minimal change in the its diffusion rate. Carbon dioxide though only has two possible orientations, and this limitation drastically reduces the diffusion rate of the gases, making it slower than both nitrogen and krypton, even though its minimum shape would suggest it would be faster than both. This points out the importance a whole range of factors has on the diffusion of the gases and that this is not simply controlled by one factor alone.

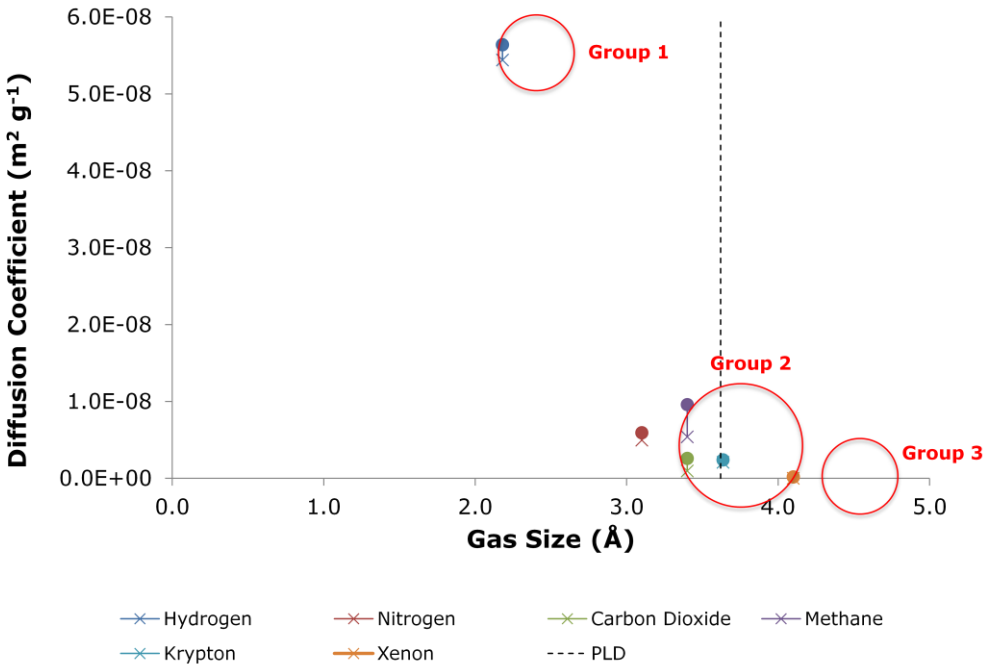


Figure 5.52 – Figure showing three clusters of gases; group 1 where the gases are much smaller than the PLD, group 2 where the gases are slightly smaller or a similar size to the PLD and group 3 where the gases are much larger than the PLD. Crosses and circles represent the RBM and FBM diffusion coefficients respectively.

5.12. Conclusion

To conclude, it has been possible to use CSFF to perform two different types of MD simulations – the RBM and the FBM. This has allowed examination into how increased flexibility in the system directly impacts diffusion rates, gas pathways and overall diffusivity. It has been shown that the flexibility of **CC3** becomes more important as the size of the gas was increased. Sulfur hexafluoride showed that not all the gases can diffuse through **CC3** and that there was some kind of threshold in gas size that limits diffusion; this was calculated to be around 4.5 Å. The increased flexibility, though, did help in rationalising the diffusion of xenon

through **CC3**. If the RBM MD simulation for xenon was considered in isolation, it would suggest that no diffusion was possible. This was because the size of xenon exceeds the PLD, therefore restricting the xenon to one cage. The increased flexibility in the FBM MD simulation for xenon allowed the PLD to periodically increase in size and this allowed xenon to slowly diffuse through the system. This helped to rationalise the uptake seen in the experimental sorption isotherms. The thermodynamic calculations showed how the well-defined adsorption sites for **CC3** increased the adsorption energies for xenon, and therefore although kinetically xenon was much slower than krypton, the gas selectivity of xenon over krypton was very high due to the strong presence of **CC3** to adsorb xenon. This would suggest that **CC3** is an ideal candidate for xenon/krypton separation.

In addition, the shape of the gas was also important. Carbon dioxide and methane showed that only specific orientations of the gases were seen to diffuse through the cage window, indicating some kind of shape exclusivity. This was possible as the shape and size of the cage windows is very close to that of the gases. By using the knowledge about how the size and shape of the gases affects diffusion through **CC3**, it could be possible to suggest ideal candidates that could be trapped within a cage – something small enough that the cages could be built around it, but too big to diffuse through the cage window. This can be achieved by comparing the size of the guests to the PLE, and subsequently selecting something that is larger than the PLE but also still fits within the cage cavity.

The PLE has unlocked a new way to screen porous materials. Instead of using the PLD, which assumes the system remain rigid, the flexibility of the system can now be included; this is far more accurate, especially when the PLD and gas size are similar. In addition, dynamic connectivity has been introduced as a valuable visualisation technique. This colour codes the surface area dependent on how long each section of the pore topology is open for, over a given period. In the future this could be used to explain the opening and closing of channels, the connection of isolated voids to the connected pore structure, as well as previously unexplainable sorption isotherms for porous materials.

Finally, no cooperative diffusion was observed for the gases and this would suggest that the motion of the cages was unaffected by the presence of other molecules within the system. This means that an empty system of **CC3** could be simulated and dynamic connectivity used to "screen" porous cage molecules to see what gases might be separated. This should give a good indication to what should and should not diffuse before running the experimental calculations. It

would be important to see whether increased loading of gases within **CC3** would affect the dynamic motion of the cages, as if this was the case this assumption would become invalid.

5.13. References

(1) Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. Nat. Mat. **2009**, *8*, 973–978.

(2) Hasell, T.; Chong, S. Y.; Jelfs, K. E.; Adams, D. J.; Cooper, A. I. J. Am. Chem. Soc. **2012**, *134*, 588–598.

(3) Sun, X.; Wick, C. D.; Thallapally, P. K.; McGrail, B. P.; Dang, L. X. J. Phys. Chem. B. **2011**, *115*, 2842–2849.

(4) Liu, J.; Keskin, S.; Sholl, D. S.; Johnson, J. K. J. Phys. Chem. C. **2011**, *115*, 12560–12566.

(5) Gurdal, Y.; Keskin, S. J. Phys. Chem. C. **2013**, 117, 5229–5241.

(6) Gupta, K. M.; Chen, Y.; Jiang, J. J. Phys. Chem. C. **2013**, *117*, 5792-5799.

(7) Cabrales-Navarro, F. A.; Gómez-Ballesteros, J. L.; Balbuena, P. B. J. Mem. Sci. **2013**, *428*, 241–250.

(8) Haldoupis, E.; Nair, S.; Sholl, D. S. J. Am. Chem. Soc. **2010**, *132*, 7528–7539.

(9) Firouzi, M.; Wilcox, J. Microporous Mesoporous Mat. **2012**, *158*, 195-203.

(10) Vela, S.; Huarte-Larrañaga, F. Carbon. **2011**, *49*, 4544–4553.

(11) Adisa, O. O.; Cox, B. J.; Hill, J. M. Eur. Phys. J. B. **2010**, *79*, 177–184.

(12) Larsen, G. S.; Lin, P.; Siperstein, F. R.; Colina, C. M. Adsorption. **2010**, *17*, 21–26.

(13) Jones, J. T. A.; Hasell, T.; Wu, X.; Bacsa, J.; Jelfs, K. E.; Schmidtmann, M.; Chong, S. Y.; Adams, D. J.; Trewin, A.; Schiffman, F.; Cora, F.; Ben Slater; Steiner, A.; Day, G. M.; Cooper, A. I. Nature. **2011**, *474*, 367–371.

(14) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Microporous Mesoporous Mater. **2012**, *149*, 134–141.

(15) Sudik, A. C.; Millward, A. R.; Ockwig, N. W.; Côté, A. P.; Kim, J.; Yaghi, O. M. J. Am. Chem. Soc. **2005**, *127*, 7110–7118.

(16) Fernandez, C. A.; Liu, J.; Thallapally, P. K.; Strachan, D. M. J. Am. Chem. Soc. **2012**, *134*, 9046−9049.

(17) Fujita, M.; Umemoto, K.; Yoshizawa, M.; Fujita, N. Chem. Commun.

2001, 509–518.

(18) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L.; Snurr, R. Q. Chem. Sci. **2012**, *3*, 2217.

(19) http://www.ccdc.cam.ac.uk/products/csd/radii.

(20) Amirjalayer, S.; Tafipolsky, M. Angew. Chem. **2006**, *46*, 463 –466.

(21) Seehamart, K.; Nanok, T.; Krishna, R.; Baten, J. M. V.; Remsungnen, T.; Fritzsche, S. Microporous Mesoporous Mater. **2009**, *125*, 97–100.

(22) Melchionna, S.; Ciccotti, G.; Lee Holian, B. Mol. Phys. **1993**, *78*, 533–544.

(23) Smith, W.; Forester, T. R. J. Mol. Graph. **1996**, *14*, 136–141.

(24) Accelrys; Accelrys Software, Inc.: San Diego, **2011**.

(25) Cöté, A. P. Science. **2005**, *310*, 1166–1170.

(26) Kärger, J.; Ruthven, D. M. *Diffusion in Zeolites and other Microporous Solids*, Wiley, **1992**.

(27) Skoulidas, A. I.; Sholl, D. S. J. Phys. Chem. A. **2003**, *107*, 10132–10141.

(28) Horike, S.; Shimomura, S.; Kitagawa, S. Nat. Chem. **2009**, *1*, 695–704.

(29) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. Nature. **2003**, *423*, 705–714.

(30) Haldoupis, E.; Nair, S.; Sholl, D. S. J. Am. Chem. Soc. **2012**, *134*, 4313–4323.

(31) Ananchenko, G. S.; Moudrakovski, I. L.; Coleman, A. W.; Ripmeester, J. A. Angew. Chem. Int. Ed. Engl. **2008**, *47*, 5616–5618.

(32) Cooper, A. I. Adv. Mater. 2009, 21, 1291–1295.

(33) Lloyd, G. O.; Barbour, L. J. Thesis: University of Stellenbosch. Dept. of Chemistry; Science, P. **2006**.

(34) Barbour, L. J. Chem. Commun. **2006**, 1163-1168.

(35) Talu, O.; Myers, A. L. Colloids Surf. A. **2001**, *187-188*, 83-93.

(36) Hasell, T.; Schmidtmann, M. J. Am. Chem. Soc. **2011**, *133*, 14920–14923.

(37) Li, W.; Grimme, S.; Krieg, H.; Möllmann, J.; Zhang, J. J. Phys. Chem. C. **2012**, *116*, 8865–8871.

Chapter 6

Conclusion and future work

6.1. Conclusion

The aim of this thesis was to accurately simulate the dynamic nature of known porous organic cage molecules, with a view to simulating the diffusion of different gases through their pore structures. Initial simulations showed that, due to their unique chemical structure, no 'off-the-shelf' force field (FF) was accurate enough to describe their dynamic motion. Therefore, a cage specific force field (CSFF) was developed to be transferable across the first three cage systems, **CC1**-**CC3**. This encompassed the intramolecular FF terms required to describe the cage structures, along with the intermolecular interactions between the molecules that determine the crystal packing. CSFF was initially based on the polymer consistent force field (PCFF), with any missing internal degrees of freedom fitted using a combination of X-ray crystal structure data and 6-31G(d,p) DFT calculations as reference data. The main adaptation was in the parameters required to describe the imine groups conjugated to an aromatic ring, as this parameter set was absent in PCFF. The intermolecular interactions between the discrete cage units were also altered, to scale the dispersion to an appropriate level. This scaling was then checked against MD simulations to show that the lattice parameters of the cage crystal structures were reproduced. Upon completion, CSFF accurately reproduced gas phase models of single **CC1**, **CC2** and **CC3** molecules as well as their crystalline solid state structures. CSFF was also transferable to other related imine cage systems, for example **CC4**, and a co-crystal of two cages (**CC1**, **CC3**). In addition to fitting the missing internal degrees of freedom for the cages, the parameters for seven different gases were incorporated. This made it possible to simulate the diffusion of these gases through the pore structure, to provide insight into separation capabilities.

Other members of our group have used CSFF for MD simulation studies.[1,2] In addition, CSFF has been used to rationalise the 'on'/'off' porosity observed in two different polymorphs of **CC1**. A combination of computational simulations, including simulated surface area calculations, GCMC adsorption isotherms, and MD simulations, for hydrogen and nitrogen in **CC1α** and **CC1β**, helped to confirm experimental results, as well as to provide further insight into why the polymorphism of **CC1** alters the porosity of the molecule. It was evident that each molecule of **CC1** possessed an internal cage cavity, and this was represented by a tetrahedron. Both hydrogen and nitrogen are small enough to diffuse into this, via the four access points, the cage windows, and as a result the packing motif of each structure is vital. The relative alignment of different cage windows in the crystal packing dictates the accessibility to each cage. For the

CC1 polymorphs to be porous to either gas, a fully connected channel through the crystal packing of the structure is required.

The computationally determined pore connectivity for **CC1α** showed that the cages packed in such a way that the cavities remain in isolation from one another (Figure 6.1a). MD simulations also verified that the lack of pore connectivity prohibited diffusion of either gas from cage to cage, although sorption into these cavities was observed in the GCMC simulations; therefore agreeing with the minimal uptake seen experimentally. The simulations of **CC1β** agreed with the experimental findings in that it is possible to crystallise discrete units of **CC1**, such that the material was selectively porous to hydrogen over nitrogen. A small decrease in the gas size (from a radius of 1.55 to 1.09 Å) was enough to switch the computationally determined pore connectivity of the crystal from being porous to hydrogen (Figure 6.1b) to being non-porous to nitrogen (Figure 6.1c). This subtle change was further validated by MD simulations, which showed that hydrogen is free to diffuse throughout the whole system, while nitrogen is confined to a single cage.

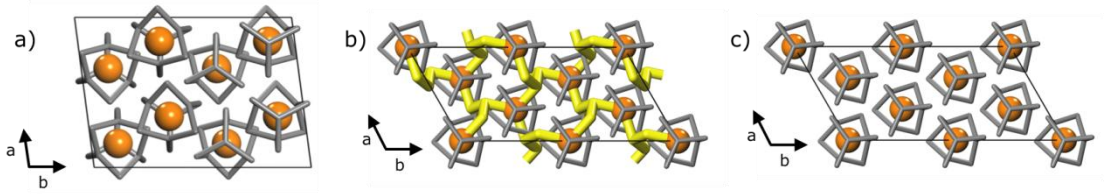


Figure 6.1 – Computationally determined pore connectivity for both **CC1** polymorphs. a) Hydrogen and nitrogen in **CC1α** show only isolated voids (orange), while, b) the isolated voids in **CC1β** are connected for hydrogen, with the layers being connected via an extrinsic pore pathway, while c) shows that only isolated voids (orange) are present in the pore structure of **CC1β** for nitrogen.

One issue encountered during this work was that the GCMC calculations carried out on both **CC1** polymorphs consistently overestimated the gas uptake. Overprediction of gas uptake in GCMC simulations is a common problem for porous materials and has been well documented in the literature for MOFs.[3-7] This could be down to a number of factors. First, although it is possible to crystallise **CC1** in a non-porous state to both hydrogen and nitrogen, each cage still has an inbuilt cavity. The simulations load gas into these cavities, thus generating results that contradict experimental observations, as there is no consideration of a connectivity pathway for them. To combat this, it would be beneficial to add a criterion to the GCMC simulations that first checked that specific sites were connected to the pore structure before permitting gas sorption. One way to do this would be to run an *a priori* MD simulation – this could then be used to map

out the pore topology of a system and highlight only the accessible sites for use in the GCMC simulations. This would automatically rule out any sorption in **CC1α** and nitrogen sorption in **CC1β**, thus reproducing experimental findings.

Overestimation of hydrogen in **CC1β** was also observed, and this again a common problem in porous materials.[7] This was rationalised either low crystallinity of the sample producing an artificially low gas uptake or,[8] more likely, the FF used, CSFF, insufficiently describing the interactions between the cages and the gases. This is unsurprising, as CSFF was not parameterised for the gas-cage interactions; this would be prudent if GCMC adsorption isotherms were to be simulated in the future.

In addition, CSFF was used to study the diffusion of a range of gases through crystalline **CC3**-*R*. Seven gases were chosen: hydrogen, nitrogen, carbon dioxide, methane, sulfur hexafluoride, krypton and xenon. These were simulated using two types of MD simulations – the rigid body method (RBM) and the flexible body method (FBM). This allowed the impact that host flexibility had on gas diffusion to be quantified, while gas pathways were also examined. The diffusion abilities of the gases are summarised in Figure 6.2.

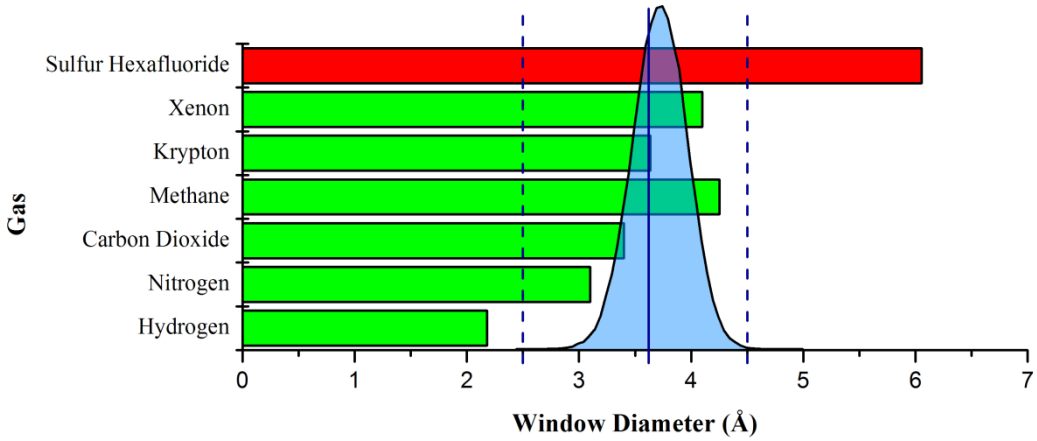


Figure 6.2 – Graph showing the pore limiting envelope (PLE) of **CC3** versus the size of the different gases. It was clear that opposed to a single value for the window diameter (solid line), the PLD, the PLE generates a minimum and maximum (dashed lines). Green bars indicate the gases that would be expected to diffuse through the system; red bars indicate the gases restricted by the window diameter.

The RBM MD simulations, restricted by the pore limiting diameter (PLD), permitted diffusion of hydrogen, nitrogen, carbon dioxide, methane and krypton, whilst the FBM MD simulations also facilitated diffusion of xenon. The PLE has highlighted an alternatives way to screen porous molecular materials. Instead of using the PLD, which assumes he system remain rigid, the flexibility of the

system can now be included; this is far more accurate, especially when the PLD and gas size are similar. The addition of larger sulfur hexafluoride in this study showed that there was a threshold to what size gases could diffuse through **CC3**. Subsequent studies on the PLE of the other gases indicated that this threshold was around 4.5 Å.

My research has also led to new methodologies and techniques being devised, and these could be applied to other systems in the future. One such analysis technique is dynamic connectivity.* One limitation in using 'static' X-ray crystal structure to describe the pore connectivity of a system is that this assumes that the host material remains fixed. This assumption may be valid for porous materials that exhibit large pore channels, for example some large pore MOFs. However, if the pore channels are a similar size to the guests, the flexibility of the system may have a direct impact on guest diffusivity, and therefore to ignore its flexibility is unreasonable. Such phenomena as cooperative diffusion, breathing modes, vibrational frequencies, the formation of transient open channels, as well as diffusion of gas into isolated voids, could be missed; to address this, dynamic connectivity should be considered. Dynamic connectivity is the examination of how the surface area of a porous system evolves over a specific time period. A MD simulation is undertaken and subsequently the pore topology of each frame is generated – these were then superimposed to calculate the pore connectivity over a given time period.

Throughout this study, no cooperative diffusion was observed for **CC3**, and this would suggest that the motion of the cages was unaffected by the gases. Cooperative diffusion is when the gas only moves through a system when the host structure accommodates such a movement by changing shape or orientation in response to the gas. If no cooperative diffusion is observed, it means that an empty system of **CC3** could be used to simulate the dynamic connectivity for a number of gases, just by altering the probe radius used. This data could then be used to 'screen' against gas size, and this should give a good indication as to what would and would not be expected diffuse through **CC3**. Any advantageous properties predicted could then be measured experimentally; one such example is the predicted separation of krypton from xenon in **CC3**. These noble gases have attracted a large amount of interest industrially, due to their physiochemical properties.[9] The relative size of krypton and xenon to the PLD of **CC3** suggested that krypton should diffuse through the system, while xenon should be restricted

* This work has been carried out in collaboration with Maciej Haranczyk.

to a single cage; this concept was shown in the RBM MD simulations. The FBM MD simulations, however, had better agreement with the measured uptakes, showing that although slow, xenon was able to diffuse through **CC3**. This meant that by allowing the flexibility allowed in the MD simulations, the PLD was able to periodically increase in size, thus allowing xenon to diffuse slowly through the system. This highlighted that the flexibility of **CC3** became an increasingly crucial factor as the size of the PLD and the gas became similar. Since experimental adsorption isotherms show that xenon is adsorbed in **CC3**, this work has shown the importance of describing **CC3** as a flexible system in simulations. The speed at which krypton and xenon diffused through **CC3** varied greatly, and diffusion coefficients of 2.4 x 10^{-9} m^2 s^{-1} and 1.83 x 10^{-10} m^2 s^{-1} were measured for the respective gases, suggesting that kinetic separation could be feasible. In addition to this, thermodynamic calculations showed that **CC3** had well-defined adsorption sites for xenon that increased the adsorption energies, while no such sites were observed for krypton. Therefore, although xenon was kinetically much slower than krypton, the preference of xenon adsorption over krypton was very high due to the strong preference of **CC3** to adsorb xenon. The rationale for this is that, compared to krypton, xenon is the 'perfect' fit, and it has the 'snuggest' fit inside the cage cavity. That said, xenon is slightly smaller than the cage cavity, meaning that a slightly larger gas, for example radon, might have an even higher binding energy. The computational simulations in this work would suggest that **CC3** is an ideal candidate for xenon/krypton separation and has led to the undertaking of experimental breakthrough measurements of krypton and xenon in **CC3**; initial results look promising and these are shown in Figure 6.3. It is important to note that predictions were carried out *a priori*, with the calculations inspiring the breakthrough experiments.

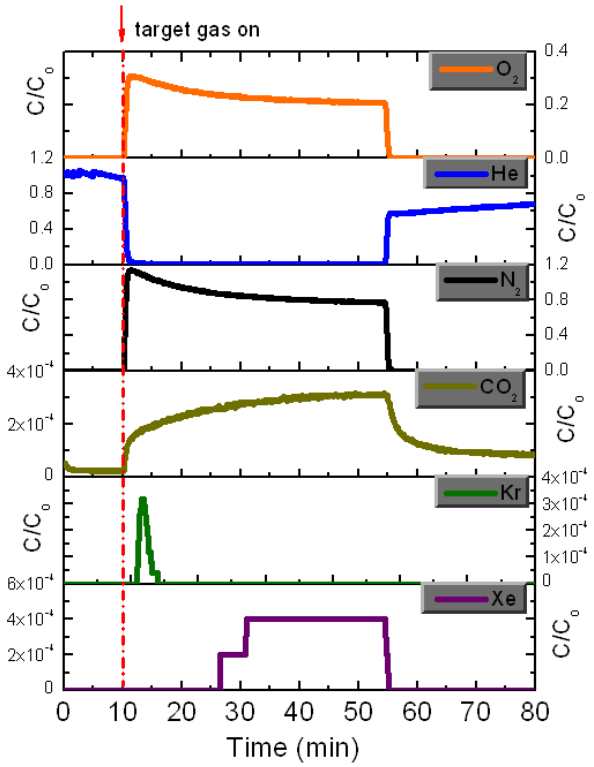


Figure 6.3 – Initial breakthrough measurements for various gases in **CC3**, carried out by Paul Reiss, showing that **CC3** has a large Kr/Xe selectivity.

In summary, the development of CSFF has made it possible to simulate the diffusion of small gases through porous organic imine cages, and it as been shown that this diffusion is dependent on the relative size of the gas to the cage window, assuming that there is a suitable diffusion pathway. Using MD simulations, we have unlocked phenomena such as gas selectivity, rare-event hopping and the diffusion of gases to regions previously thought inaccessible. This has aided the rationalisation of existing experimental observations and is a significant step forward for *a priori* prediction of porous organic cage systems properties in the future. This work has also led to directly new experiments, prompted by my simulations.

In addition, a new way to visualise the connectivity of a system has been introduced. This is achieved by monitoring how the surface area evolves with respect to time, as the MD simulation progresses.[10] This suggests how the pore channels of various systems, previously thought too small for gas adsorption, are actually ideal candidates for separations.

6.2. Future work

In the future it will be possible to use CSFF to rationalise the uptake of gases and also larger guests, such as halogens and solvents, within a range of cage systems. The concept here is to use MD analysis for *in silico* screening of cage materials, in particular molecular separations. If reliable, this would be faster than the associated sorption experiments, with calculations taking only a matter of hours. If adsorption simulations in porous organic cages were to be carried out in the future, it might be prudent to tailor the interactions between the gases and the cages within CSFF, as to date these have not been appropriately parameterised. This would increase the accuracy of the results and may provide additional information about adsorption energies and potential binding sites.

One way to analyse the pore topology of a system is to visualise dynamic connectivity. This concept was used in Chapter 5 to study the diffusivity of a number of gases in **CC3**, but can be extended to other cage systems, for example **CC2**. **CC2** was synthesised via a cycloimination reaction of 1,3,5-triformylbenzene with 1,2-propylenediamine. Visualisation of the 'static' crystal structure shows that an extrinsic one-dimensional channel runs through the system, with each cage possessing its own internal cavity (Figure 6.4a). Measured adsorption isotherms showed that **CC2** exhibited uptakes of hydrogen, nitrogen, carbon dioxide and methane, though these uptakes were too high to be rationalised solely by the one-dimensional channel. One hypothesis is that gas molecules somehow diffuses from the one-dimensional channel into the isolated voids; this is based upon an electron density map generated from single crystal X-ray diffraction data showing disordered physisorbed water within the isolated cage cavities (Figure 6.4b).[11] From a modelling perspective, MD simulations could be used to visualise the dynamic connectivity of the system for a given probe radius, to see whether connection of the one-dimensional channel and isolated voids was observed at any point. To test this, a 2 x 2 x 2 supercell of **CC2** (using the 2220 isomer) was generated and a 1 ns FBM MD simulation was undertaken at 298 K, 1 atm. Preliminary calculations showed that when the dynamic connectivity of the MD simulation was visualised, using a probe radius of 1.09 Å to represent hydrogen, connections between the one-dimensional channel and the isolated voids were observed (Figure 6.4c). This suggests that an in-depth study would be beneficial to see whether diffusion of gas into the isolated voids is observed. The positional disorder of the methyl group on the diamine cage vertex further complicates this study, as this may influence the pore connectivity of **CC2** and therefore the influence of their placement would also needs to be assessed.

There is the intriguing possibility that this disorder plays a key role in the accessibility of the cage cavities.

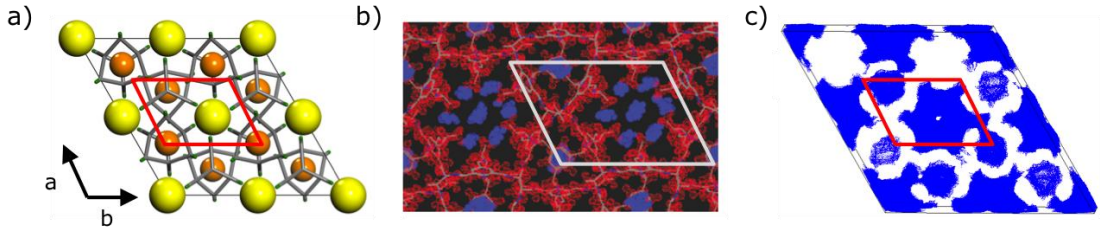


Figure 6.4 – a) Computationally determined pore connectivity for **CC2**, showing an extrinsic one-dimensional channel (yellow) and isolated cage voids (orange), b) an electron density map generated from single crystal X-ray diffraction data showing disordered physisorbed water within the isolated cage cavities, compared to, b) a simulated dynamic connectivity map for **CC2**, over a 1 ns simulations, using a hydrogen probe radius (1.09 Å).* c) This shows how the channels become connected over the simulation. A common unit cell has been highlighted on each picture.

Another avenue that could be explored is the hypothetical screening of porous organic cages. As mentioned in Chapter 4, one advantage of these systems is their processability as well as their propensity to crystallise into different polymorphs. This opens up the potential for screening crystal structure predictions (CSP), and work has already published in this area.[12] In principle, CSP could be used to generate a whole subset of hypothetical cage polymorphs, and these structures could then undergo high throughput *in silico* computational screening, based on work outlined in Chapter 5. When combined with the use of the PLE as a selection criterion, the dynamic pore connectivity, sorption and diffusion capabilities of these hypothetical targets could be used to identify ideal candidates for specific applications, such as diffusivity, encapsulation, selectivity, storage and selectivity – these could then be targeted by synthesis. Although similar work has been carried out by Snurr *et al.* for hypothetical metal organic frameworks (MOFs), they keep the host framework fixed, and therefore flexibility is not taken into consideration;[13] this assumption would be invalid for porous organic cages, as MD simulations on both **CC1** and **CC3** during this work have shown the importance of including host flexibility. Also, there is no consideration as to whether the 137,000 hypothetical MOFs proposed by Snurr *et al.* are feasible synthetic targets, since no energies were calculated. Contrary to this, CSP would allow the use of calculated lattice energies to rank the energetic feasibility of a specific polymorph, thus highlighting chemically reasonably structures that also display advantageous properties.

* This work has been carried out in collaboration with Maciej Haranczyk.

6.3. References

(1) Jelfs, K. E.; Wu, X.; Schmidtmann, M.; Jones, J. T. A.; Warren, J. E.; Adams, D. J.; Cooper, A. I. Angew. Chem. Int. Ed. **2011**, *50*, 10653–10656.

(2) Mitra, T.; Jelfs, K. E.; Schmidtmann, M.; Ahmed, A.; Chong, S. Y.; Adams, D. J.; Cooper, A. I. Nat. Chem. **2013**, *5*, 276–28.

(3) Parkes, M. V.; Staiger, C. L.; Perry, J. J., IV; Allendorf, M. D.; Greathouse, J. A. Phys. Chem. Chem. Phys. **2013**, *15*, 9093.

(4) Thornton, A. W.; Furman, S. A.; Nairn, K. M.; Hill, A. J.; Hill, J. M.; Hill, M. R. Microporous and Mesoporous Mater. **2013**, *167*, 188–197.

(5) Dubbeldam, D.; Frost, H.; Walton, K. S.; Snurr, R. Q. Fluid Phase Equilib. **2007**, *261*, 152–161.

(6) Düren, T.; Bae, Y.-S.; Snurr, R. Q. Chem. Soc. Rev. **2009**, *38*, 1237–1247.

(7) Li, W.; Grimme, S.; Krieg, H.; Möllmann, J.; Zhang, J. J. Phys. Chem. C. **2012**, *116*, 8865–8871.

(8) Hasell, T.; Zhang, H.; Cooper, A. I. Adv. Mater. **2012**, *24*, 5732–5737.

(9) Dorcheh, A. S.; Denysenko, D.; Volkmer, D.; Donner, W.; Hirscher, M. Microporous and Mesoporous Mater. **2012**, *162*, 64–68.

(10) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Microporous and Mesoporous Mater. **2012**, *149*, 134–141.

(11) Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. Nat. Mat. **2009**, *8*, 973–978.

(12) Jones, J. T. A.; Hasell, T.; Wu, X.; Bacsa, J.; Jelfs, K. E.; Schmidtmann, M.; Chong, S. Y.; Adams, D. J.; Trewin, A.; Schiffman, F.; Cora, F.; Ben Slater; Steiner, A.; Day, G. M.; Cooper, A. I. Nature. **2011**, *474*, 367–371.

(13) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L.; Snurr, R. Q. Chem. Sci. **2012**, *3*, 2217.

Appendix

All supplementary information

A.1. Positional isomerism for CC2

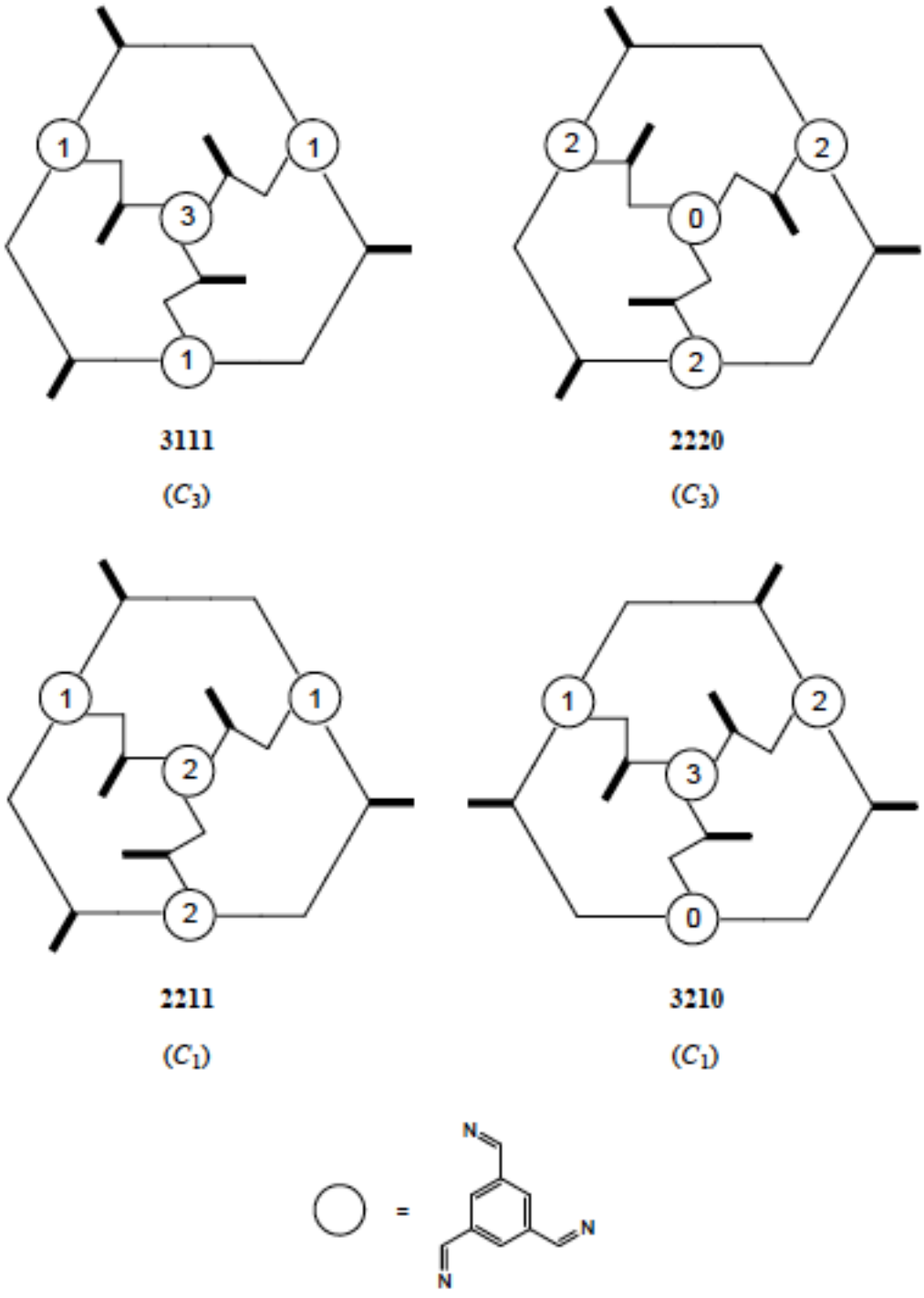


Figure S1 – This figure shows the **CC2** positional isomerism. The methyl group on the vertices are disordered over the *exo*-positions, which gives rise to four isomers, which corresponds to eight enantiomers, as the cage structure is chiral. These are shown above: **3111**, **2220**, **2111**, and **3210**. These numbers indicate how many methyl groups are adjacent to each aromatic ring. Only **2220** is considered during this work. Figure taken from reference [1].

A.2. Fitting the bond parameters

Bond stretch potentials are fitted using the quartic potential defined in equation 2.31 in the main text. The fitted parameters are compared to the standard PCFF values in Table S5.

Graph	i	J	r_0 (Å)	k_2	k_3	k_4
S1	cp	cp	1.3838	465.2720	-1230.7532	1460.3640
PCFF	cp	cp	1.4170	470.8361	-627.6179	1327.6345
S2	cp	h	1.0787	421.9058	-1038.8437	1199.9151
PCFF	cp	h	1.0982	372.8251	-803.4526	894.3173
S3	c=1	cp	1.4556	353.8001	-741.6000	936.9211
PCFF	c=1	cp	1.5000	322.8000		
S4	c=1	h	1.1005	365.7679	-725.5404	781.6621
PCFF	c=1	h	1.0900	361.6000		
S5	c=1	n=	1.2700	758.1000	-1675.0000	2153.0000
PCFF	c=1	n=	1.4100	331.8740		
S6	c	n=	1.4329	324.7000	-218.0000	3030.0000
PCFF	c	n=	1.4750	336.8000		
S7	c	h	1.1010	345.0000	-691.8900	844.6000
PCFF	c	h	1.1010	345.0000	-691.8900	844.6000
S8	c	c	1.5360	250.2000	-503.2000	400.9000
PCFF	c	c	1.5300	299.6700	-501.7700	679.8100

Table S1. Comparison of the new bond potential parameters fitted as part of CSFF and the original PCFF values.

Note: Where no k_3 and k_4 values are given for the PCFF reference potentials only a quadratic form of is available.

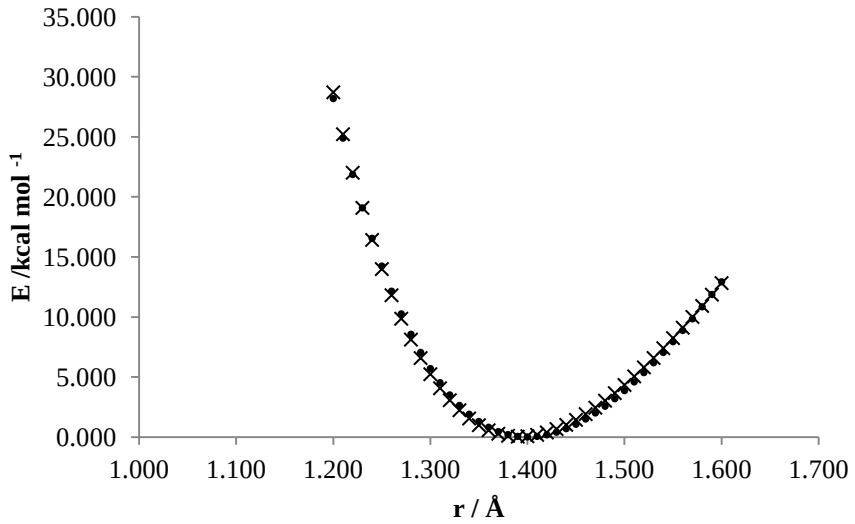


Figure S2. Energy, E vs atom atom separation, r, for the cp..cp bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S1.

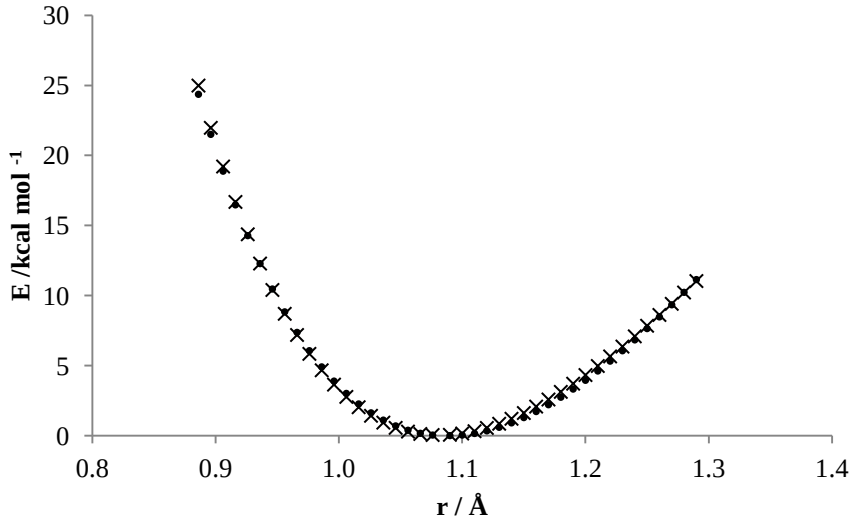


Figure S3. Energy, E vs atom atom separation, r, for the cp..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S1.

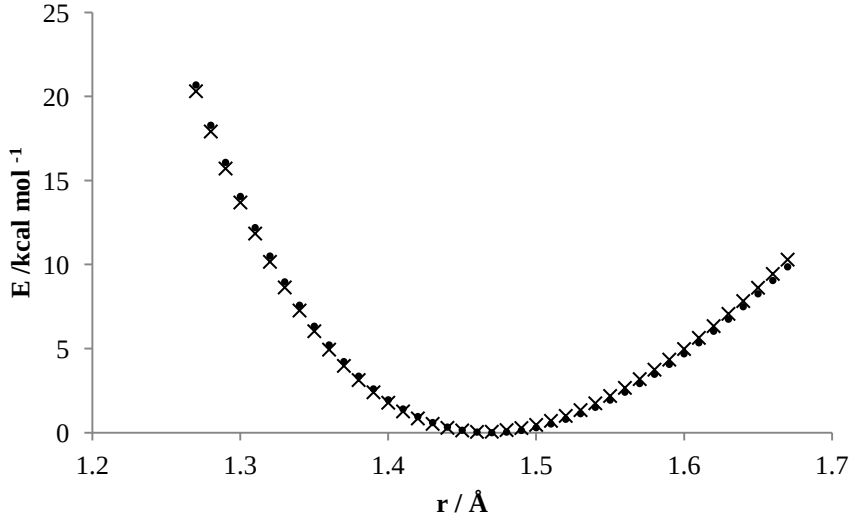


Figure S4. Energy, E vs atom atom separation, r, for the cp..c=1 bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S1.

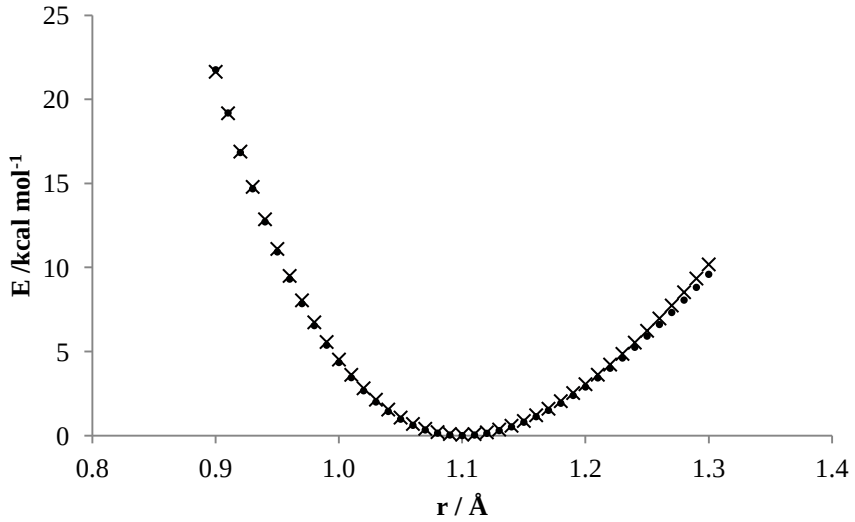


Figure S5. Energy, E vs atom atom separation, r, for the c=1..hc bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S1.

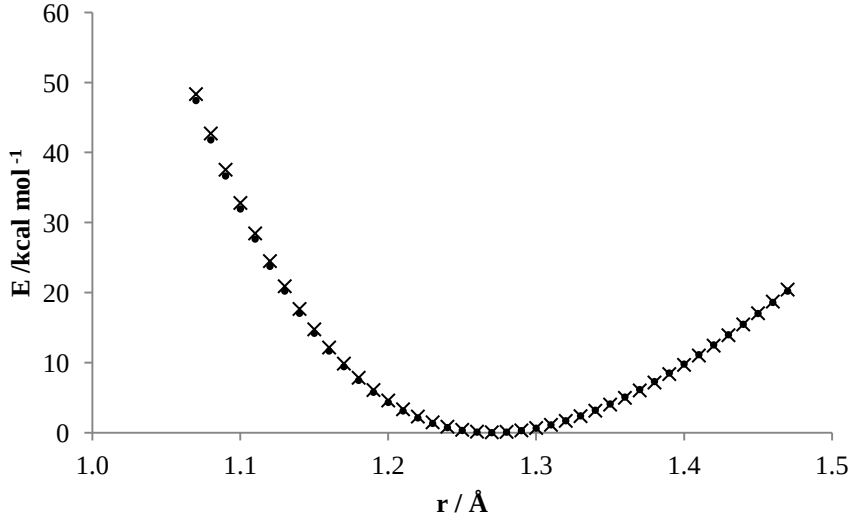


Figure S6. Energy, E vs atom atom separation, r, for the c=1..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S1.

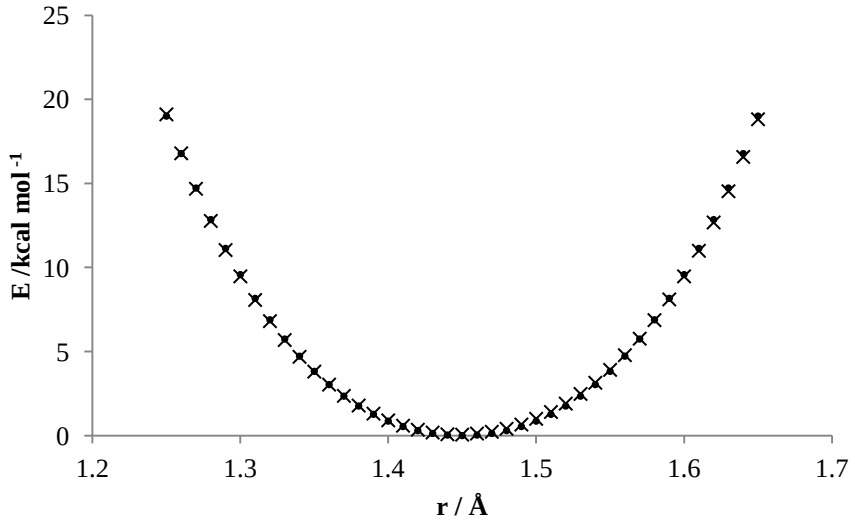


Figure S7. Energy, E vs atom atom separation, r, for the c..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S1.

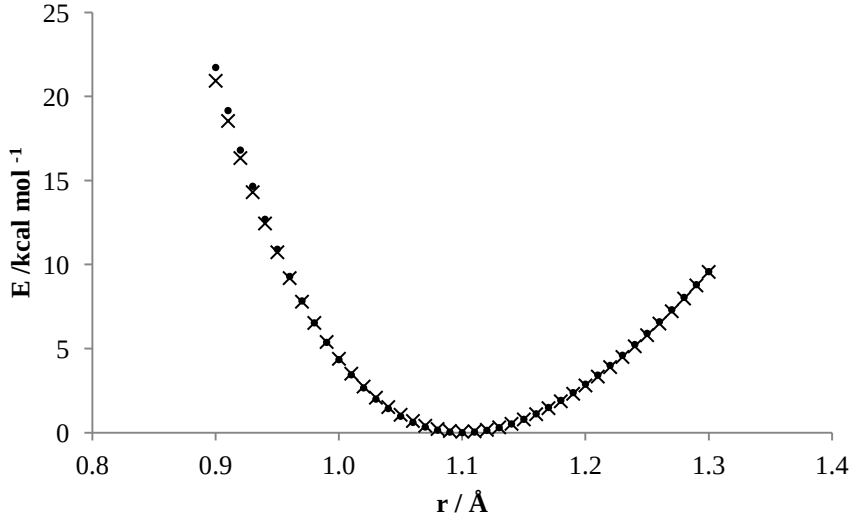


Figure S8. Energy, E vs atom atom separation, r, for the c..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S1.

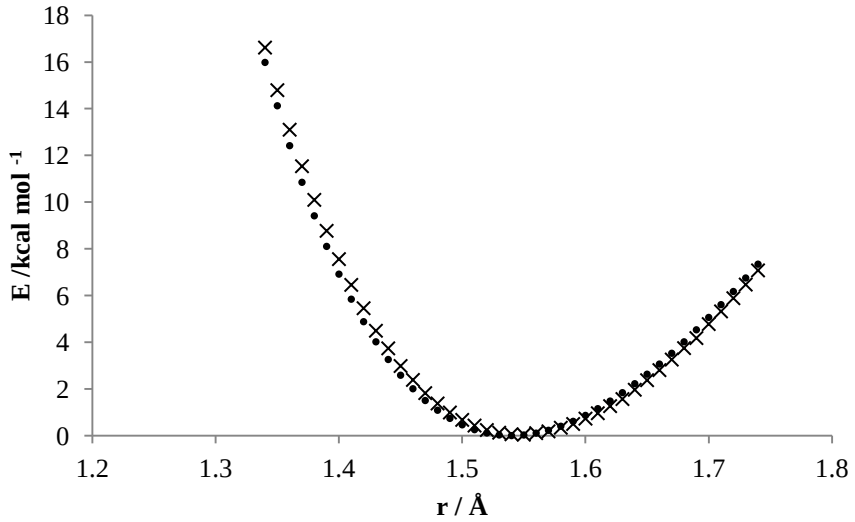


Figure S9. Energy, E vs atom atom separation, r, for the c..c bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S1.

A.3. Fitting the angle parameters

Angle potentials are fitted using the quartic potential defined in equation 2.34 in the main text. The fitted parameters are compared to the standard PCFF values in Table S6.

Graph	i	j	k	θ_0 /°	k_2	k_3	k_4
S9	cp	cp	h	117.9400	35.1558	-12.4682	0.0000
PCFF	cp	cp	h	117.9400	35.1558	-12.4682	0.0000
S10	cp	cp	cp	119.9000	62.0226	-0.9931	0.0000
PCFF	cp	cp	cp	118.9000	61.0226	-34.9931	0.0000
S11	c=	cp	cp	115.4201	37.1311	0.6510	1.3200
PCFF	c=	cp	cp	120.0000	65.0000		
S12	cp	c=	h	117.4110	31.5039	-11.1174	-10.4170
PCFF	cp	c=	h	120.0000	37.5000		
S13	cp	c=	n=	123.6700	57.0000	-20.7468	38.2873
PCFF	cp	c=	n=	120.0000	90.0000		
S14	h	c=	n=	122.0000	34.0000	-9.0000	-15.0000
PCFF	h	c=	n=	120.0000	37.5000		
S15	c=	n=	c	110.5000	57.9000	-67.0001	51.7001
PCFF	c=	n=	c	120.0000	60.0000		
S16	c	c	n=	110.0047	53.4431	-59.0000	0.0000
PCFF	c	c	n=	117.2847	55.4431	0.0000	0.0000
S17	h	c	n=	109.7000	36.0247	1.0032	0.0000
PCFF	h	c	n=	107.4989	62.7484	0.0000	0.0000
S18	c	c	h	109.7700	37.4530	-10.6040	5.1290
PCFF	c	c	h	110.7700	41.4530	-10.6040	5.1290

Table S2. List of all the angle potential parameters input into CSFF.

Note: Where no k_3 and k_4 values are given for the PCFF reference potentials only a quadratic form of is available.

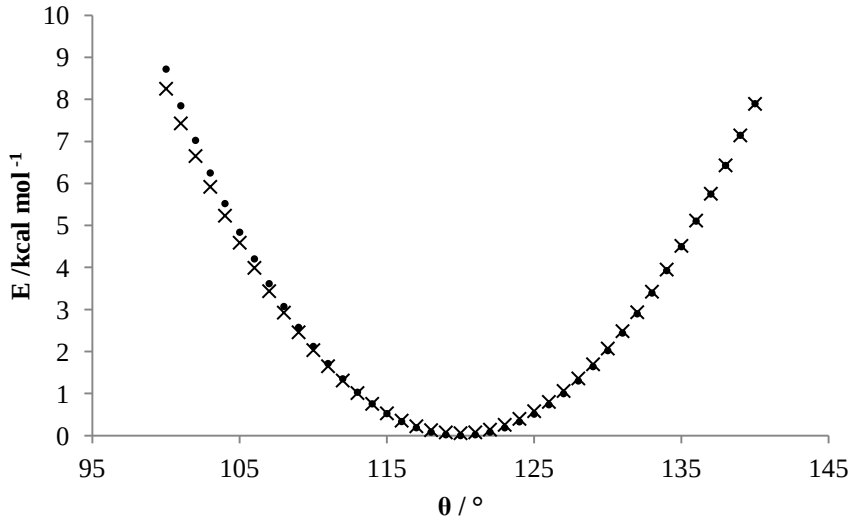


Figure S10. Energy, E vs angle, θ, for the cp..cp..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

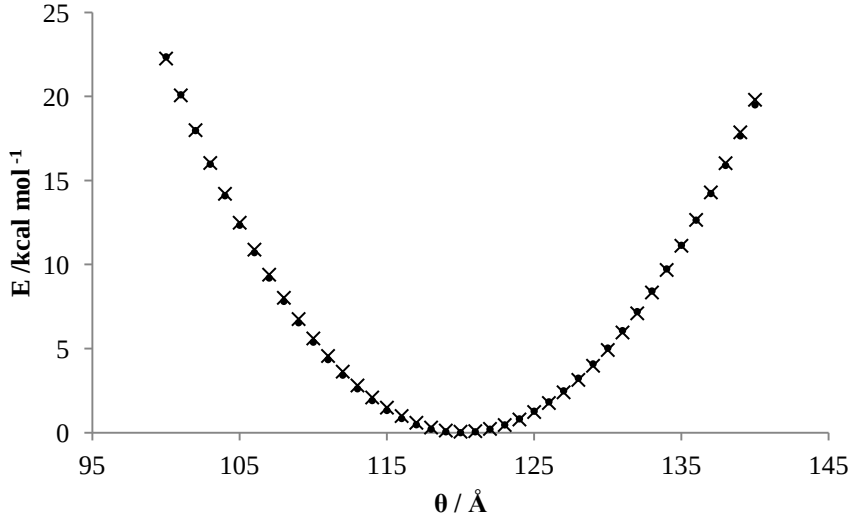


Figure S11. Energy, E vs angle, θ, for the cp..cp..cp bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

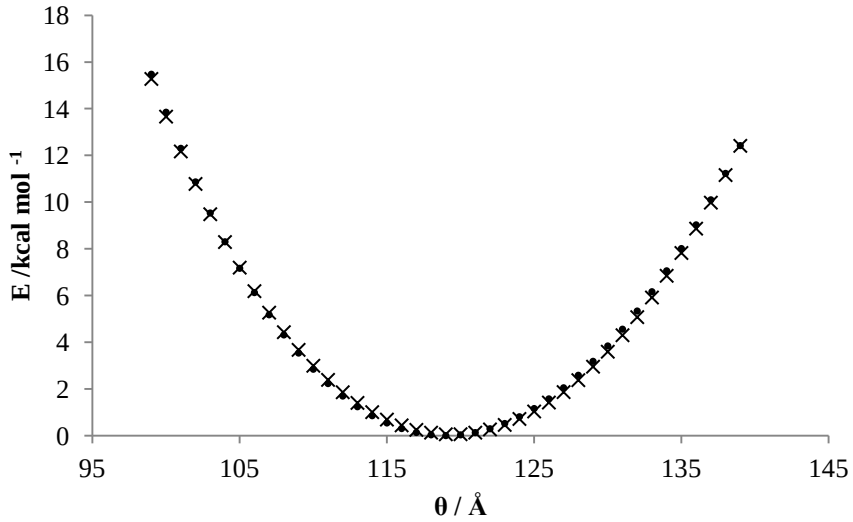


Figure S12. Energy, E vs angle, θ, for the c=..cp..cp bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

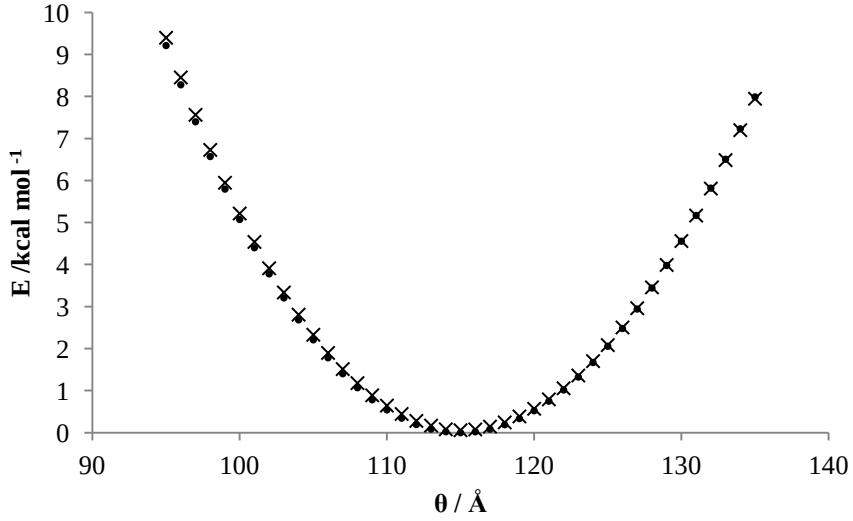


Figure S13. Energy, E vs angle, θ, for the cp..c=..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

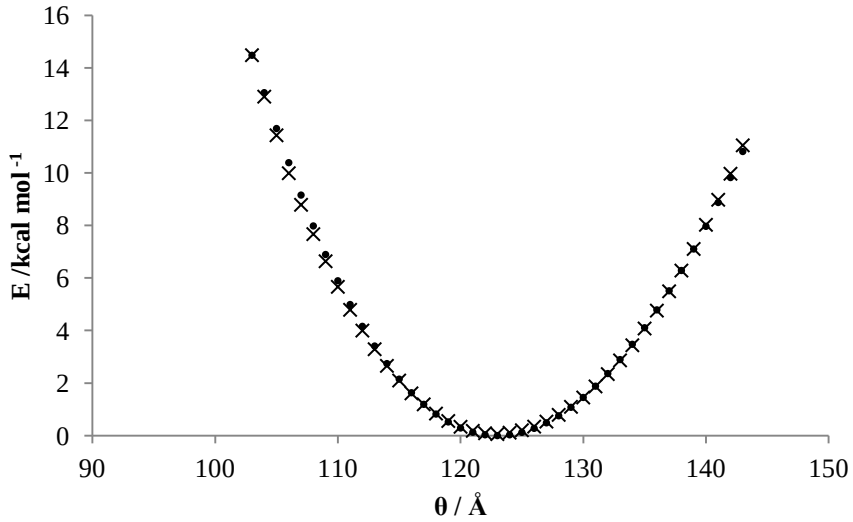


Figure S14. Energy, E vs angle, θ, for the cp..c=..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

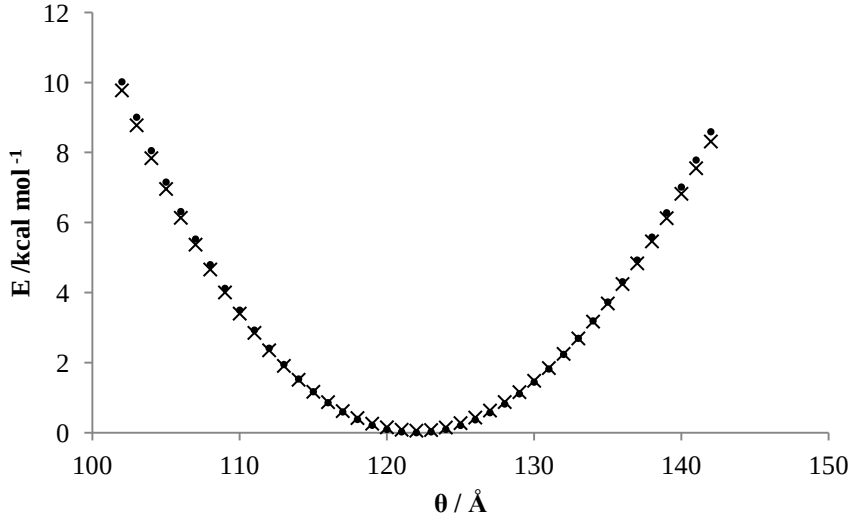


Figure S15. Energy, E vs angle, θ, for the h..c=..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

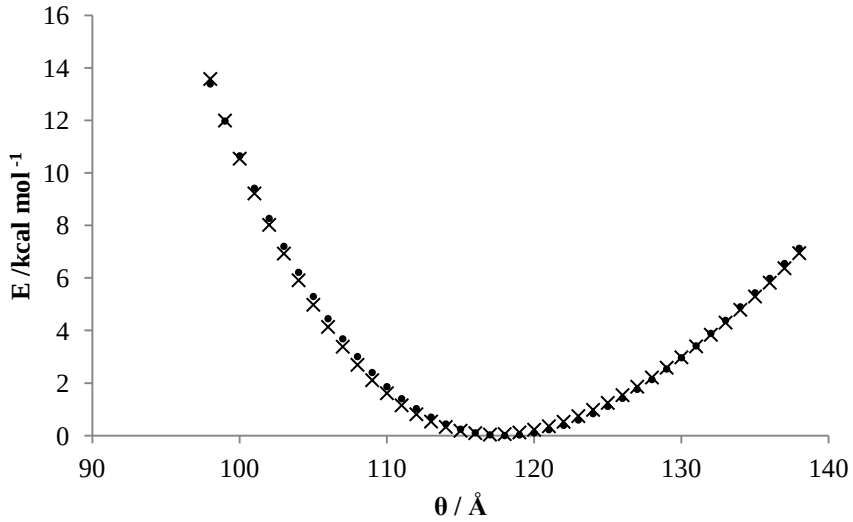


Figure S16. Energy, E vs angle, θ, for the c=..n=..c bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

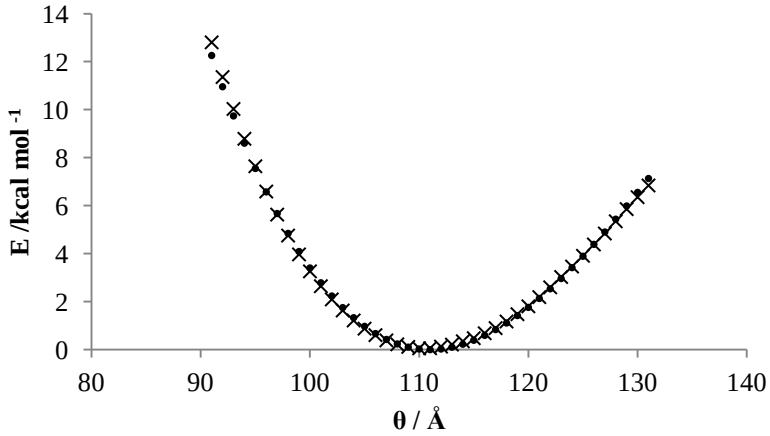


Figure S17. Energy, E vs angle, θ, for the c..c..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

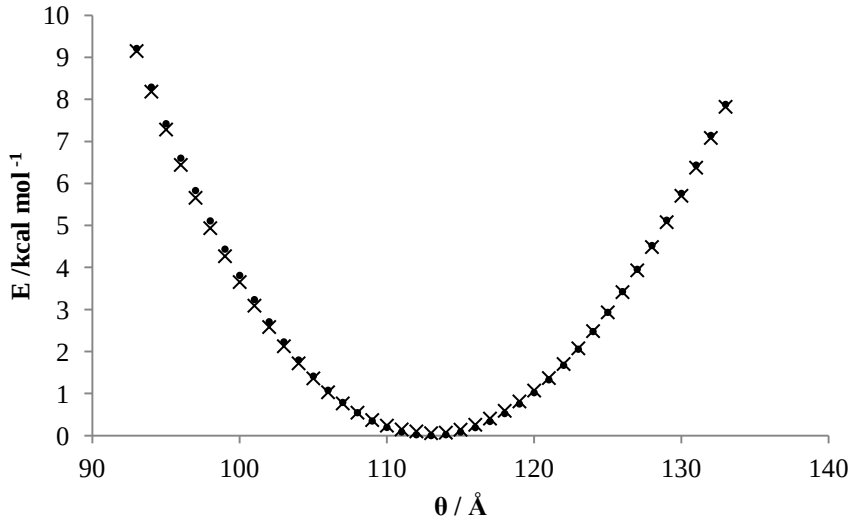


Figure S18. Energy, E vs angle, θ, for the h..c..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

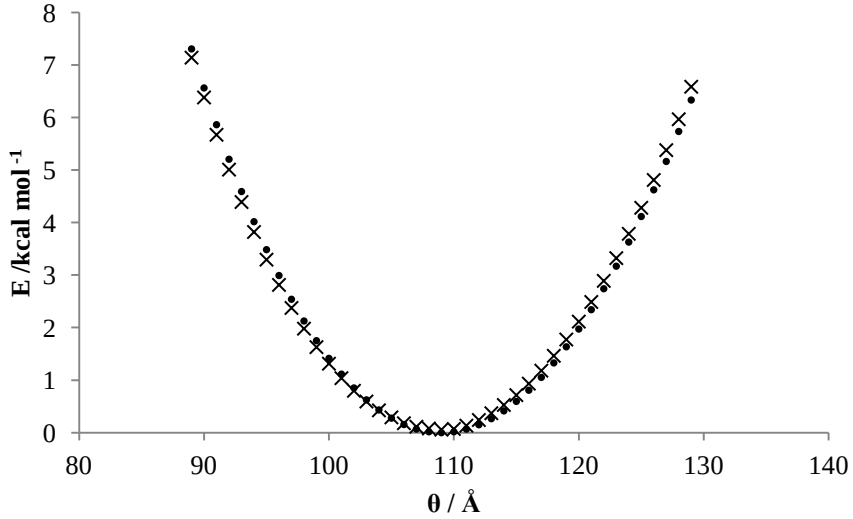


Figure S19. Energy, E vs angle, θ, for the c..c..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S2.

A.4. Fitting the dihedral parameters

Dihedral potentials are fitted using the potential defined in equation 3 of the main text. The fitted parameters are compared to the standard PCFF values in Table S7.

Graph	i	j	k	l	V_1	V_2	V_3
S19	h	cp	cp	h	4.5000	1.8769	0.0000
PCFF	h	cp	cp	h	0.0000	1.8769	0.0000
S20	cp	cp	cp	h	-7.0000	2.0661	0.0000
PCFF	cp	cp	cp	h	0.0000	3.9661	0.0000
S21	cp	cp	cp	cp	8.3667	4.1932	0.0000
PCFF	cp	cp	cp	cp	8.3667	1.1931	0.0000
S22	c=1	cp	cp	cp	-3.0000	3.3072	0.9700
PCFF	c=1	cp	cp	cp			
S23	c=1	cp	cp	h	7.0000	2.5072	0.0000
PCFF	c=1	cp	cp	h			
S24	h	c=1	cp	cp	1.0000	1.1097	0.0000
PCFF	h	c=1	cp	cp			
S25	cp	cp	c=1	n=	0.0000	1.0000	0.0000
PCFF	cp	cp	c=1	n=			
S26	c	n=	c=1	cp	-1.0000	7.9000	-0.8000
PCFF	c	n=	c=1	cp			
S27	c=1	n=	c	h	1.4300	-1.1200	0.8629
PCFF	c=1	n=	c	h			
S28	c=1	n=	c	c	0.8000	0.0400	0.2000
PCFF	c=1	n=	c	c			
S29	n=	c	c	n=	0.5071	0.2349	-0.2349
PCFF	n=	c	c	n=			
S30	h	c	c	n=	-0.0228	0.0280	-0.1863
PCFF	h	c	c	n=	-0.0228	0.0280	-0.1863
S31	h	c	c	h	-0.4432	-0.2617	-0.1283
PCFF	h	c	c	h	-0.2432	0.0617	-0.1383
S32	c	n=	c=1	h	3.0000	5.5000	0.4000
PCFF	c	n=	c=1	h			

Table S3. List of all the torsional angle potential parameters input into CSFF.

Note: There were no parameters for the imine region of the fragment within PCFF, therefore the torsions had no reference values and are blank. In addition, ϕ_1, ϕ_2, and ϕ_3 are all zero.

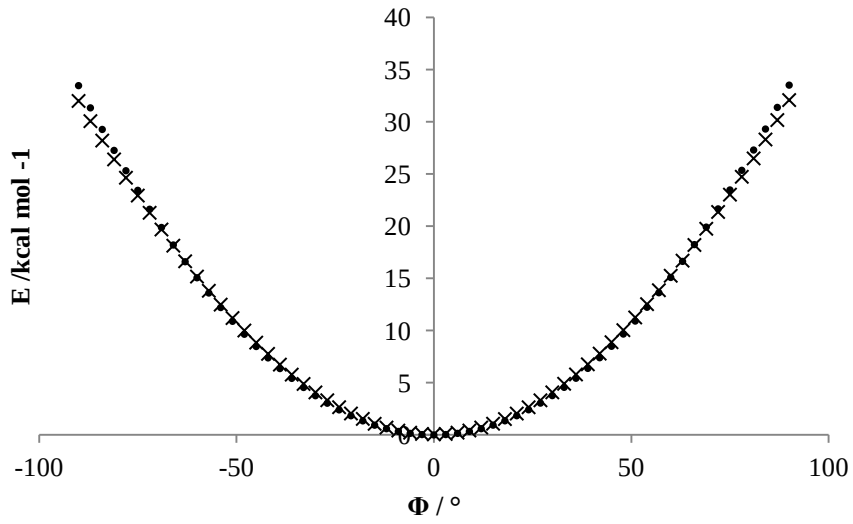


Figure S20. Energy, E vs dihedral, ϕ, for the h..cp..cp..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

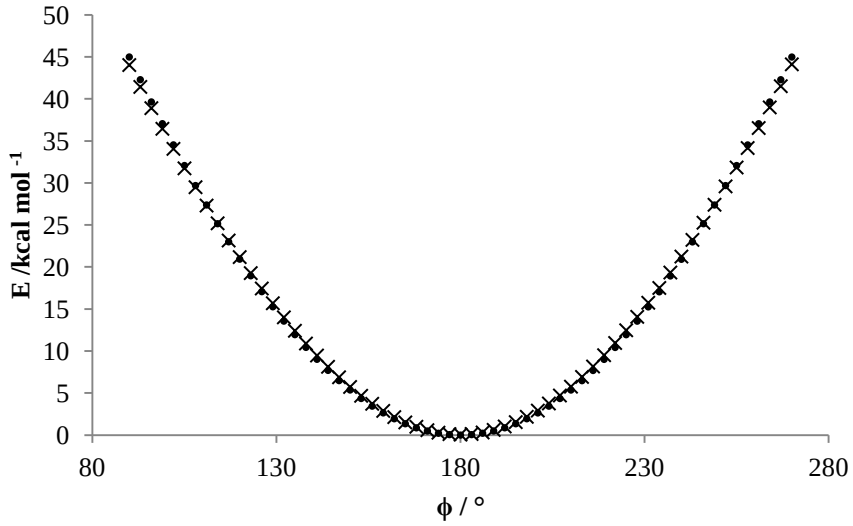


Figure S21. Energy, E dihedral, ϕ, for the cp..cp..cp..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

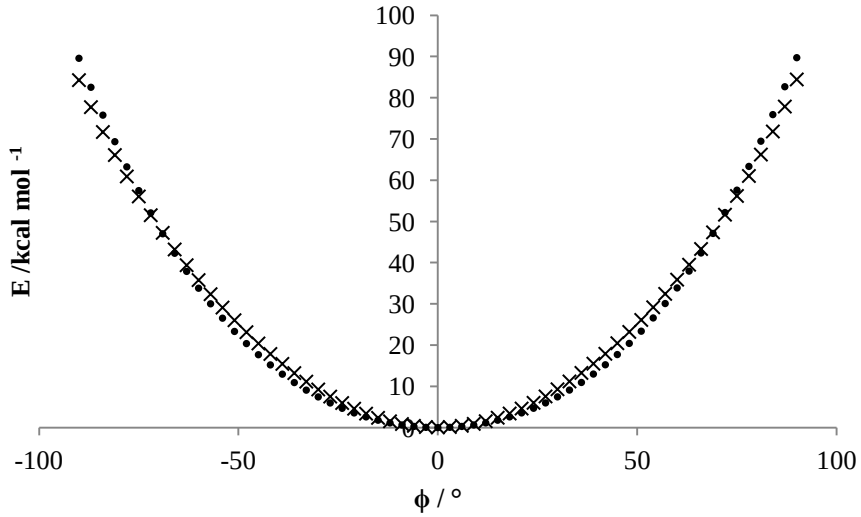


Figure S22. Energy, E dihedral, ϕ, for the cp..cp..cp..cp bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

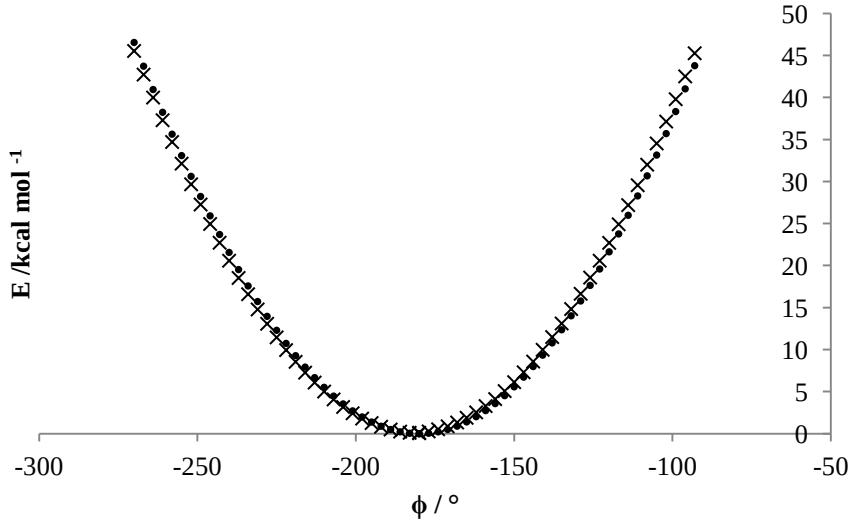


Figure S23. Energy, E vs dihedral, ϕ, for the c=1..cp..cp..cp bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

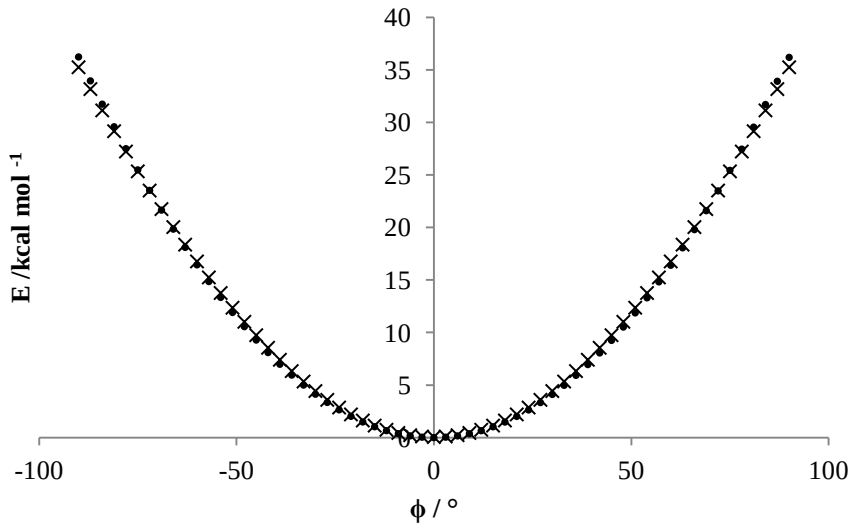


Figure S24. Energy, E vs dihedral, ϕ, for the c=1..cp..cp..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

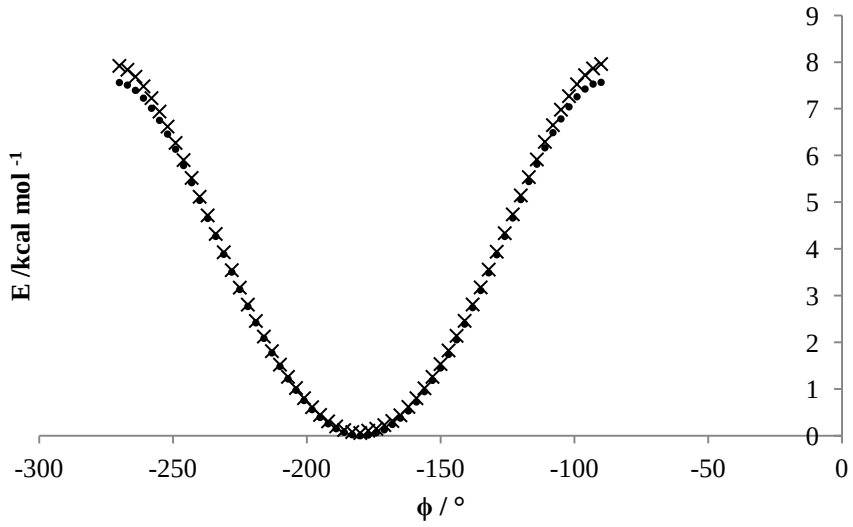


Figure S25. Energy, E vs dihedral, ϕ, for the h..c=1..cp..cp bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

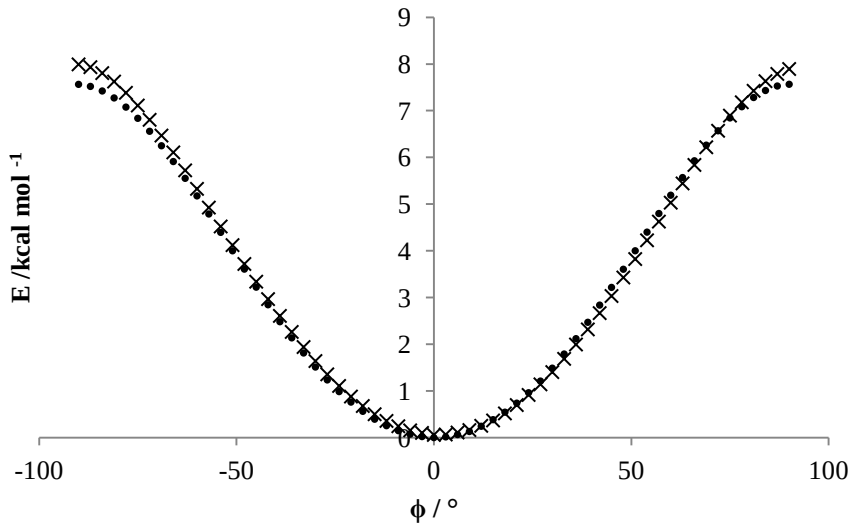


Figure S26. Energy, E vs dihedral, ϕ, for the cp..cp..c=1..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

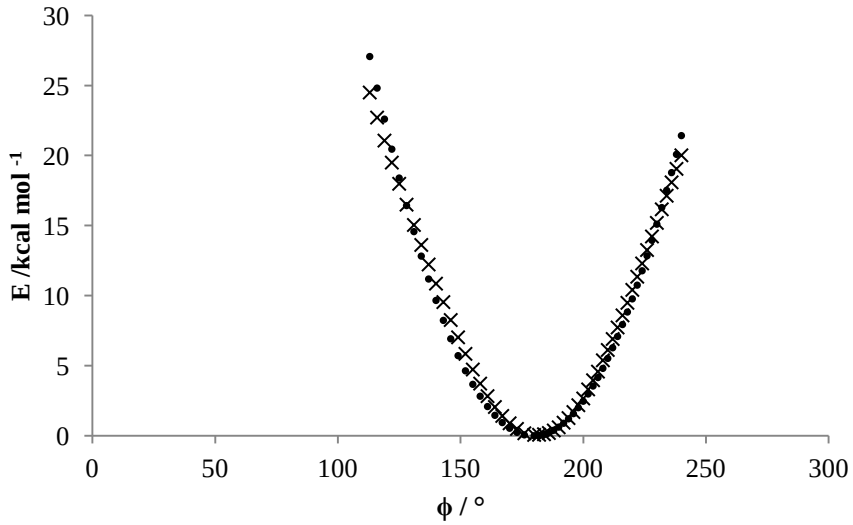


Figure S27. Energy, E vs dihedral, ϕ, for the c..n=..c=1..cp bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

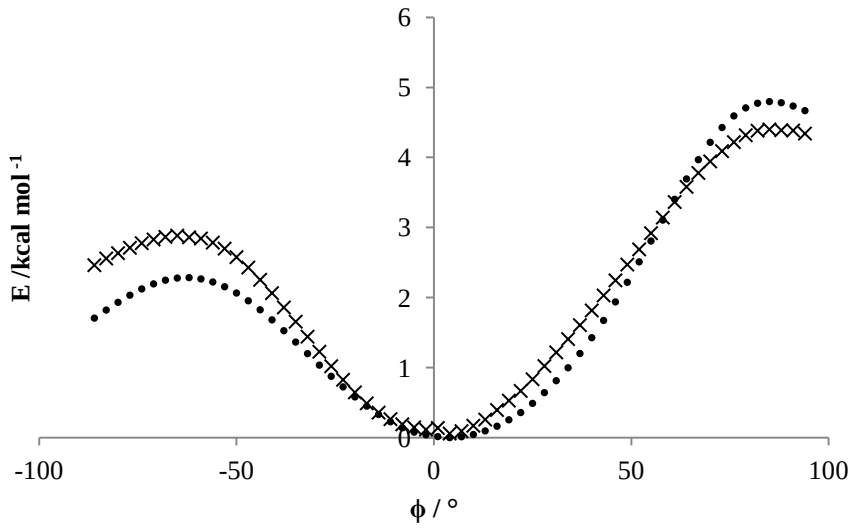


Figure S28. Energy, E vs dihedral, ϕ, for the c=1..n=..c..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

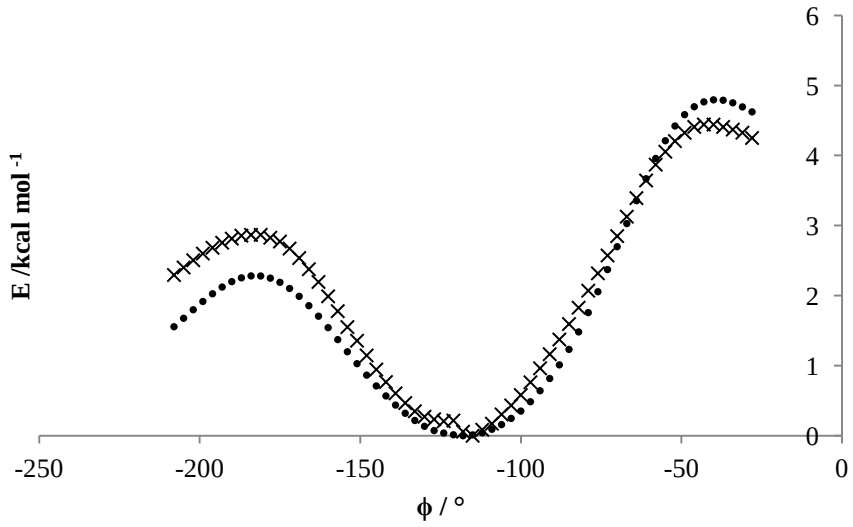


Figure S29. Energy, E vs dihedral, ϕ, for the c=1..n=..c..c bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

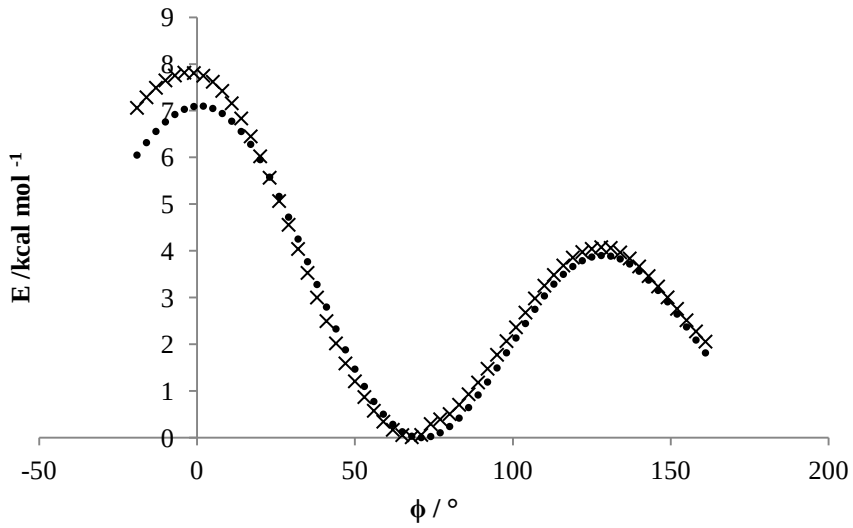


Figure S30. Energy, E vs dihedral, ϕ, for the n=..c..c..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

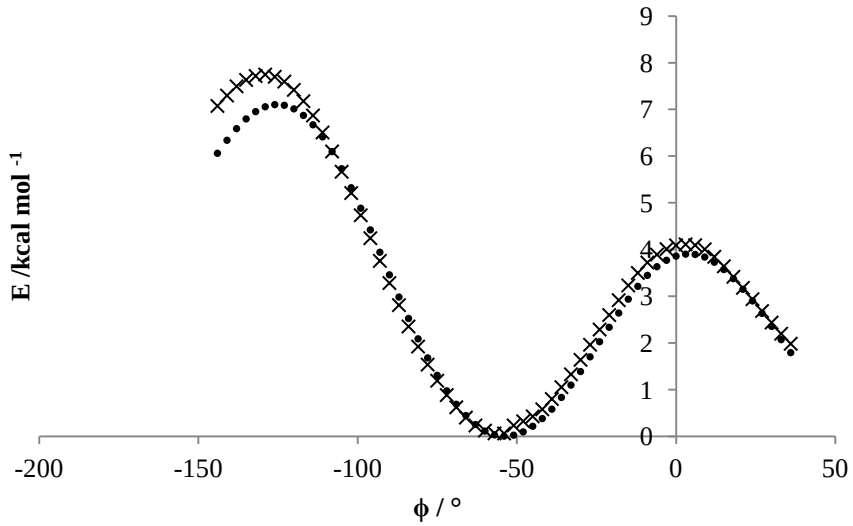


Figure S31. Energy, E vs dihedral, ϕ, for the h..c..c..n= bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

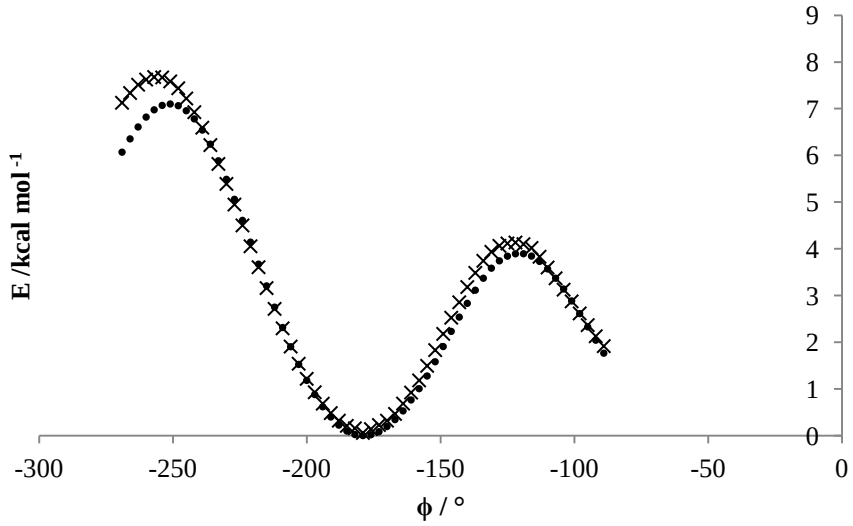


Figure S32. Energy, E vs dihedral, ϕ, for the h..c..c..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

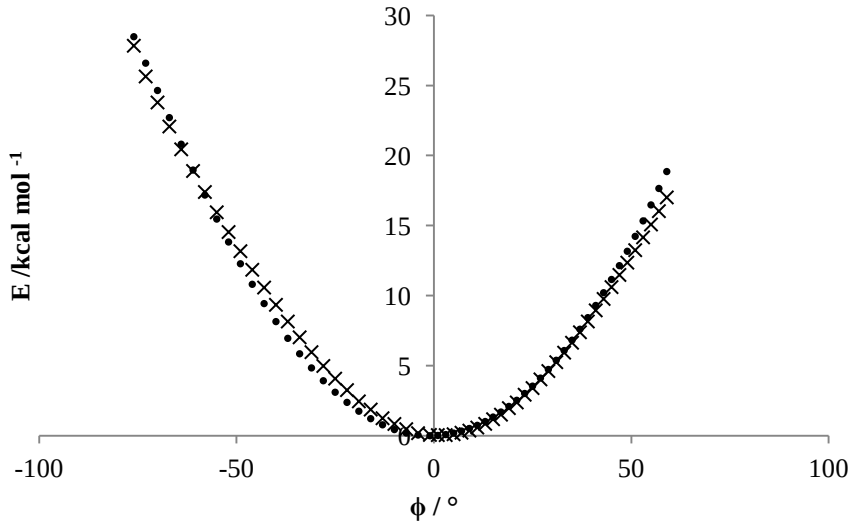


Figure S33. Energy, E vs dihedral, ϕ, for the c..n=..c=1..h bond, from the reference B3LYP/6-31G(d,p) – PCFF data for the fragment (filled circles) and the CSFF potential (crosses) using the parameters given in Table S3.

A.5. Parameterisation of the gases

Below are all the LJ potentials showing how well the potentials are matched to one another.

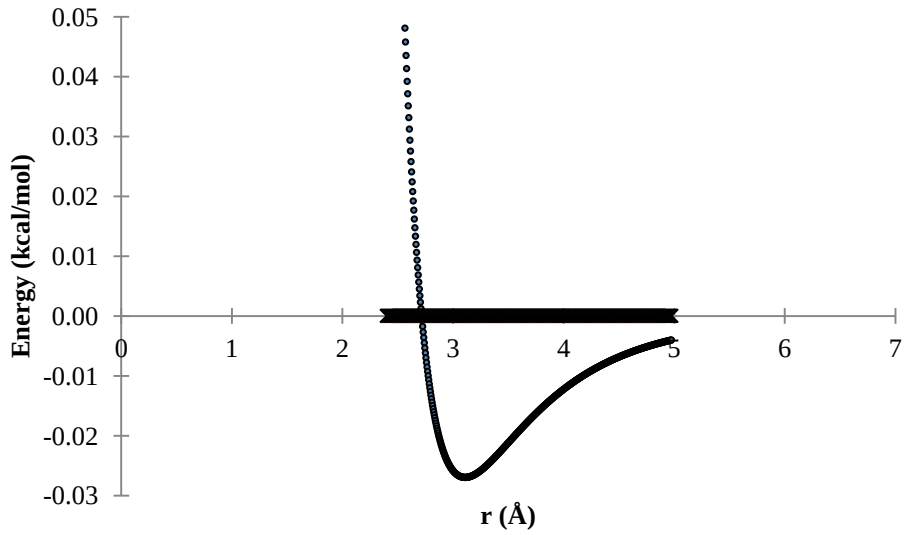


Figure S34. The energy, E vs. atomic distance, r, for a 9-6 (circles) and a 12-6 (crosses) LJ potential for HQ in hydrogen gas.

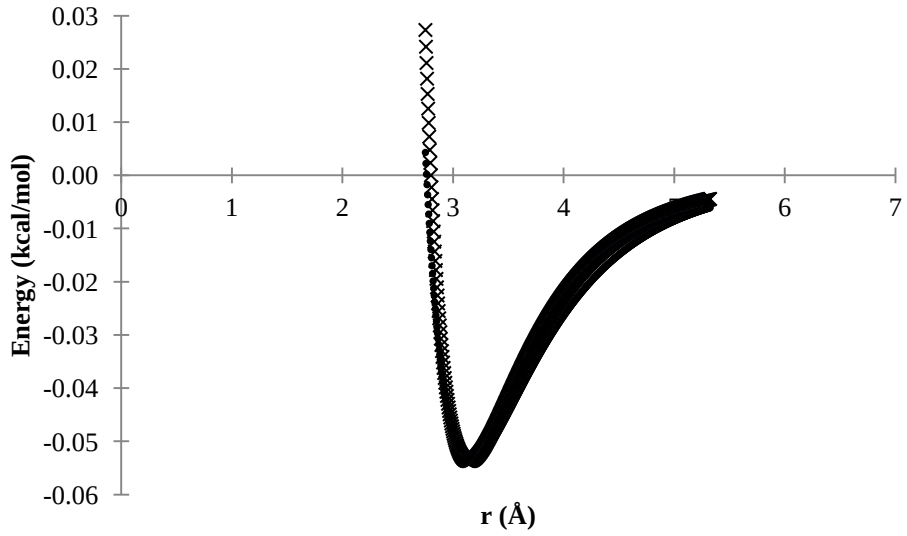


Figure S35. The energy, E vs. atomic distance, r, for a 9-6 (circles) and a 12-6 (crosses) LJ potential for CQ, in carbon dioxide.

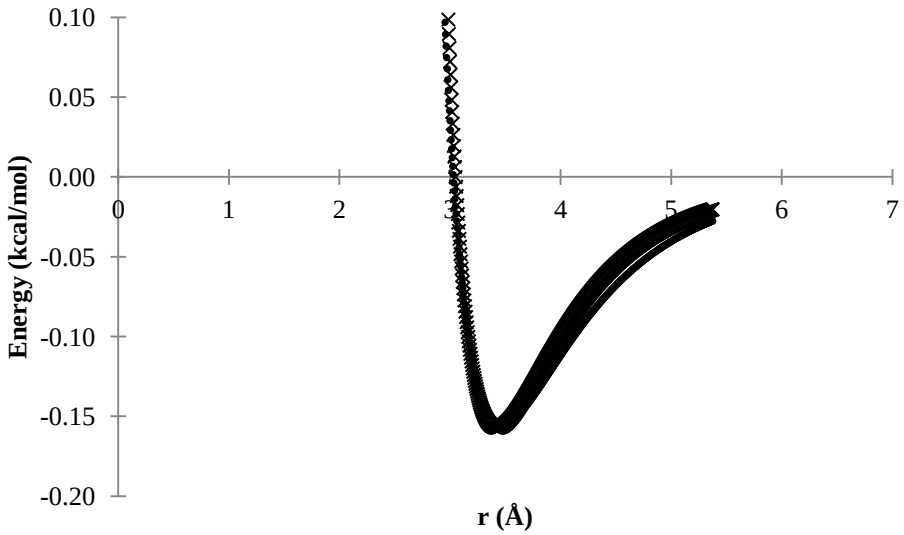


Figure S36. The energy, E vs. atomic distance, r, for a 9-6 (circles) and a 12-6 (crosses) LJ potential for OQ, in carbon dioxide.

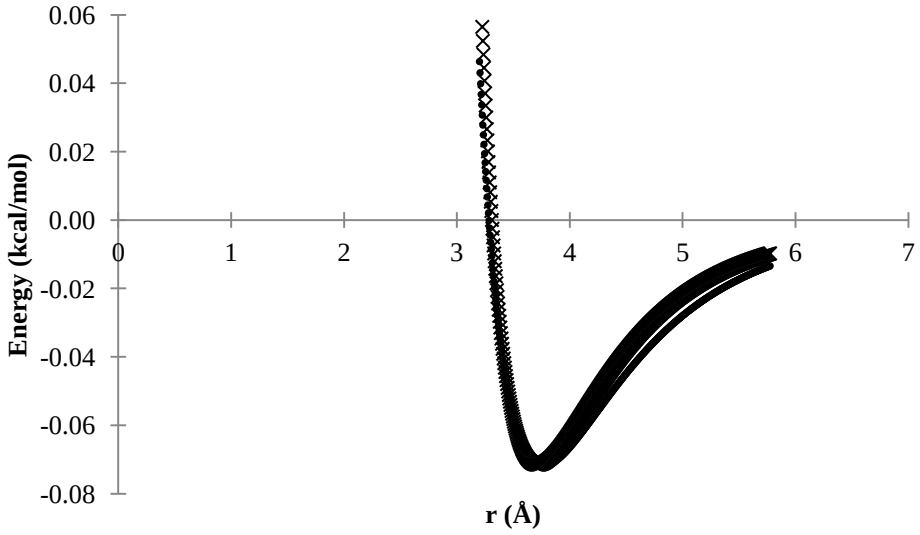


Figure S37. The energy, E vs. atomic distance, r, for a 9-6 (circles) and a 12-6 (crosses) LJ potential NQ, in nitrogen gas.

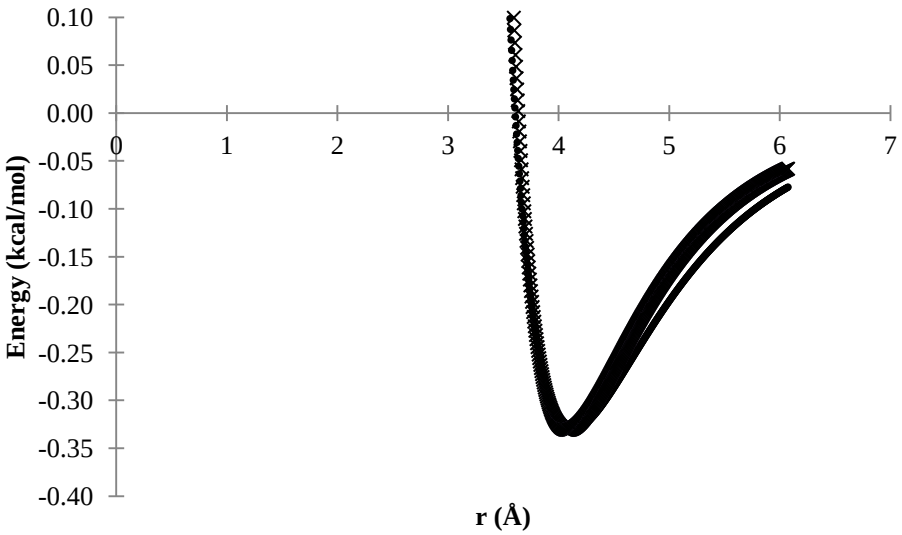


Figure S38. The energy, E vs. atomic distance, r, for a 9-6 (circles) and a 12-6 (crosses) LJ potential for krypton gas.

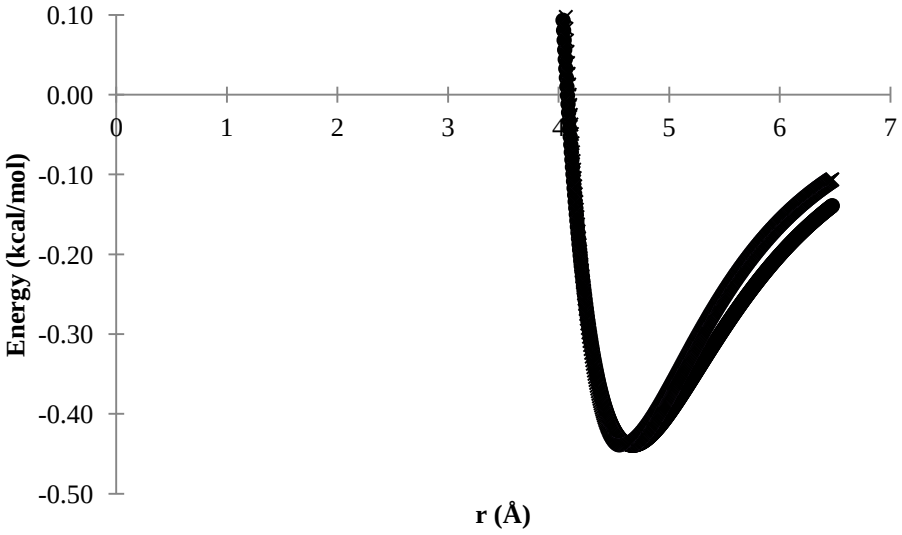


Figure S39. The energy, E vs. atomic distance, r, for a 9-6 (circles) and a 12-6 (crosses) LJ potential for xenon gas.

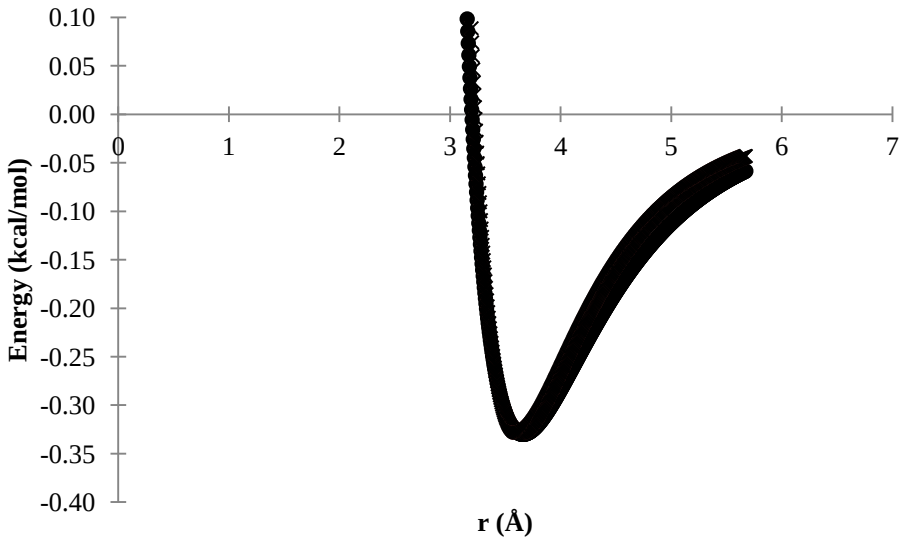


Figure S40. The energy, E vs. atomic distance, r, for a 9-6 (circles) and a 12-6 (crosses) LJ potential for SQ, in sulfur hexafluoride.

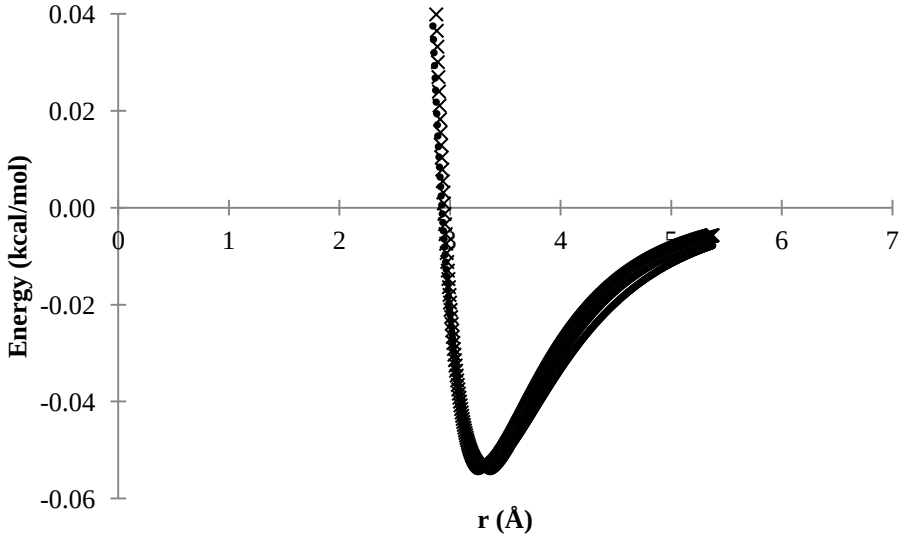


Figure S41. The energy, E vs. atomic distance, r, for a 9-6 (circles) and a 12-6 (crosses) LJ potential for FQ, in sulfur hexafluoride.

A.6. Examples from literature of diffusivity through porous materials

Material	Gas	D_s (m^2 s^{-1})	Reference
ZIF-68	Carbon Dioxide	~ $1.00x10^{-9}$	2-6
ZIF-69	Carbon Dioxide	~ $1.00x10^{-10}$	3,7
ZIF-68	Carbon Dioxide	~$1.00x10^{-10}$	8-14
ZIF-70	Carbon Dioxide	~$1.50x10^{-9}$	5,9
ZIF-68	Methane	~$1.00x10^{-9}$	9,15
ZIF-70	Methane	$1.00x10^{-8}$	9,16
ZIF-68	Hydrogen	$2.00x10^{-8}$	9,17
ZIF-70	Hydrogen	$2.00x10^{-7}$	9,18
ZIF-68	Methane	$2.00x10^{-9}$	9,19
ZIF-70	Methane	$2.00x10^{-8}$	9,20,21
MOF-5	Carbon Dioxide	$1.00x10^{-8}$	22,23
MOF-5	Hydrogen	$2.00x10^{-7}$	23,24
MOF-5	Methane	$3.00x10^{-8}$	23,25
MOF-5	Nitrogen	$3.50x10^{-8}$	23
4A Zeolite	Krypton	$3.00x10^{-9}$	20,26
4A Zeolite	Methane	$3.61x10^{-8}$	26-28
4A Zeolite	Nitrogen	$4.28x10^{-7}$	26,29
ZK4 Zeolite	Methane	$2.53x10^{-9}$	30,31
Silicate	Methane	$6.58x10^{-9}$	31,32
Silicate	Methane	$1.50x10^{-9}$	27,33
Silicate	Xenon	$4.00x10^{-10}$	27,33
Amorphous Polyethylene	Methane	$7.74x10^{-9}$	34,35
Silicate	Methane	$1.45x10^{-8}$	36,37
Silicate	Sulfur Hexafluoride	$5.32x10^{-10}$	37
Silicate	Xenon	$4.04x10^{-9}$	37

Table S4 – Table displaying diffusion coefficients of gases simulated within **CC3** for other porous materials. This allows comparison between calculated and literature data.

A.7. References

(1) Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. *Nature Materials* **2009**, *8*, 973–978.

(2) Barbour, L. J.; University of Stellenbosch. Faculty of Science. Dept. of Chemistry; Science, P. **2006**.

(3) Liu, D.; Zheng, C.; Yang, Q.; Zhong, C. *J. Phys. Chem. C* **2009**, *113*, 5004–5009.

(4) Jiang, J.-X.; Su, F.; Niu, H.; Wood, C. D.; Campbell, N. L.; Khimyak, Y. Z.; Cooper, A. I. *Chem. Commun.* **2008**, 486–488.

(5) Atwood, J. L. *Science* **2002**, *298*, 1000–1002.

(6) Atwood, J. L. *Science* **2002**, *296*, 2367–2369.

(7) Dubbeldam, D.; Krishna, R.; Snurr, R. Q. *J. Phys. Chem.* **2009**.

(8) Holst, J. R.; Trewin, A.; Cooper, A. I. *Nature Publishing Group* **2010**, *2*, 915–920.

(9) Rankin, R. B.; Liu, J.; Kulkarni, A. D.; Johnson, J. K. *J. Phys. Chem. C* **2009**, *113*, 16906–16914.

(10) Barrer, R. M.; Shanson, V. H. *J Chem Soc* **1976**.

(11) Sozzani, P.; Bracco, S.; Comotti, A.; Ferretti, L.; Simonutti, R. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 1816–1820.

(12) Dalgarno, S. J.; Thallapally, P. K.; Barbour, L. J.; Atwood, J. L. *Chem. Soc. Rev.* **2007**, *36*, 236–245.

(13) Tian, J.; Thallapally, P. K.; McGrail, B. P. *CrystEngComm* **2012**, *14*, 1909.

(14) Mastalerz, M. *Angew. Chem. Int. Ed.* **2010**.

(15) Barbour, L. J. *Chem. Commun.* **2006**, 1163.

(16) Deng, H.; Grunder, S.; Cordova, K. E.; Valente, C.; Furukawa, H.; Hmadeh, M.; Gandara, F.; Whalley, A. C.; Liu, Z.; Asahina, S.; Kazumori, H.; O'Keeffe, M.; Terasaki, O.; Stoddart, J. F.; Yaghi, O. M. *Science* **2012**, *336*, 1018–1023.

(17) Van Heest, T.; Teich-McGoldrick, S. L.; Greathouse, J. A.; Allendorf, M. D.; Sholl, D. S. *J. Phys. Chem. C* **2012**, *116*, 13183–13195.

(18) Haldoupis, E.; Nair, S.; Sholl, D. S. *J. Am. Chem. Soc.* **2012**, *134*, 4313–4323.

(19) Yang, Q.; Jobic, H.; Salles, F.; Kolokolov, D.; Guillerm, V.; Serre, C.; Maurin, G. *Chem. Eur. J.* **2011**, *17*, 8882–8889.

(20) Salles, F.; Bourrelly, S.; Jobic, H.; Devic, T.; Guillerm, V.; Llewellyn, P.; Serre, C.; Férey, G.; Maurin, G. *J. Phys. Chem. C* **2011**, *115*, 10764–

10776.

(21) Salles, F.; Jobic, H.; Devic, T.; Llewellyn, P. L.; Serre, C.; Férey, G.; Maurin, G. *ACS Nano* **2010**, *4*, 143–152.

(22) Sun, X.; Wick, C. D.; Thallapally, P. K.; McGrail, B. P.; Dang, L. X. *J. Phys. Chem. B* **2011**, *115*, 2842–2849.

(23) Skoulidas, A. I.; Sholl, D. S. *J. Phys. Chem. B* **2005**, *109*, 15760–15768.

(24) Haldoupis, E.; Nair, S.; Sholl, D. S. *J. Am. Chem. Soc.* **2010**, *132*, 7528–7539.

(25) Seehamart, K.; Nanok, T.; Krishna, R.; Baten, J. M. V.; Remsungnen, T.; Fritzsche, S. *Microporous and Mesoporous Materials* **2009**, *125*, 97–100.

(26) Sircar, S. A. M. A. L. **2006**, 1–42.

(27) June, R. L.; Bell, A. T.; Theodorou, D. N. *Journal of Physical Chemistry* **1990**.

(28) Demontis, P.; Fois, E. S.; Suffritti, G. B.; Quartieri, S. *J. Phys. Chem.* **1990**, *94*, 4329–4334.

(29) Liu, J.; Keskin, S.; Sholl, D. S.; Johnson, J. K. *J. Phys. Chem. C* **2011**, *115*, 12560–12566.

(30) Vela, S.; Huarte-Larrañaga, F. *Carbon* **2011**, *49*, 4544–4553.

(31) Demontis, P.; Suffritti, G. B. *Chemical Physics Letters* **1994**, *223*, 355–362.

(32) Xue, C.; Zhong, C. *Chinese Journal of Chemistry* **2009**.

(33) Zhao, L.; Yang, Q.; Ma, Q.; Zhong, C.; Mi, J.; Liu, D. *J Mol Model* **2011**, *17*, 227–234.

(34) Combariza, A. F.; Sastre, G.; Corma, A. *J. Phys. Chem. C* **2009**, *113*, 11246–11253.

(35) Müller-Plathe, F. *J. Chem. Phys.* **1991**, *94*, 3192.

(36) Brockway, A. M.; Schrier, J. *J. Phys. Chem. C* **2013**, *117*, 393–402.

(37) Skoulidas, A. I.; Sholl, D. S. *J. Phys. Chem. B* **2002**, *106*, 5058–5067.